

Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers*

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

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

April 16, 2008
Preliminary and Incomplete

Abstract

We consider nonparametric identification of random utility models of multinomial choice using observation of consumer choices, i.e., “micro data.” Our model of preferences is nonparametric and distribution free, nesting random coefficients discrete choice models widely used in practice with parametric functional form and distributional assumptions. We allow for choice-specific unobservables, endogenous choice characteristics, unknown heteroskedasticity, and correlated taste shocks. Under standard orthogonality, “large support,” and instrumental variable assumptions, we show identifiability of the distribution of preferences conditional on any vector of observed and unobserved characteristics. We show that key maintained hypotheses are testable and develop extensions to the case of limited support and the case in which only market-level data are available.

Keywords: discrete choice, random coefficients, differentiated products, instrumental variables

*We had helpful early conversations on this topic with Rosa Matzkin and Yuichi Kitamura. We also thank Sunyong Park and seminar participants at M.I.T. for helpful comments.

1 Introduction

We consider identification of nonparametric random utility models of multinomial choice using observation of consumer choices, i.e., “micro data.” Our model of preferences nests random coefficients discrete choice models widely used in practice with parametric functional form and distributional assumptions. However, our model is nonparametric and distribution free. It incorporates choice-specific unobservables and endogenous choice characteristics, both of which are essential to modeling demand in most settings. It also permits unknown heteroskedasticity and correlated taste shocks. Under standard orthogonality, “large support,” and instrumental variable assumptions, we show identifiability of the distribution of preferences conditional on any vector of observed and unobserved characteristics. We show that key maintained hypotheses are testable and develop extensions to the case of limited support and the case in which only market-level data are available.

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

Our analysis demonstrates that with sufficiently rich data, random utility multinomial choice models featuring unobserved characteristics are identified without the parametric assumptions used in practice – typically, linear utility with independent taste shocks (entering additively

and/or multiplicatively) drawn from parametrically specified distributions. Indeed, we provide positive identification results for a more general model of preferences than any considered previously (to our knowledge) in the econometrics or applied literatures. Our results may therefore lead to greater confidence in estimates and policy conclusions obtained using estimates of discrete choice demand models. In particular, parametric specifications used in practice can correctly be viewed as parsimonious approximations in finite samples rather than essential maintained assumptions. We view this as our primary message. However, our results also suggest that with large samples even richer specifications (parametric or nonparametric) of preferences might be considered in empirical work. and our constructive identification proofs may suggest estimation approaches.

The identifiability of random utility discrete choice models is not a new question, and our results build on two well-known ideas (we relate our results more precisely to the prior literature in section 9). The first is that of a “special regressor”—an observable (or vector of observables) with large support (e.g., Manski (1985), Matzkin (1992), Lewbel (2000)). Following standard arguments, sufficient variation in such observables enables one to “trace out” the distribution of the random component of utilities for all choices within a given choice set. The second idea is the use of variation in choice characteristics, within and across choice sets, to decompose variation in the distribution of utilities into the contributions of observed and unobserved factors (e.g., Berry (1994), Berry, Levinsohn, and Pakes (1995), Berry, Levinsohn, and Pakes (2004)). Combining and extending these ideas enables us to obtain positive results for a less restrictive nonparametric model than those considered previously. In particular, the use of within market variation in consumer attributes allows us to trace out key features of demand without having to worry about cross-market variation in unobservable product characteristics. Then, the cross-market variation in choice sets allows a general treatment of unobservables.¹ In addition, our attention to identifiability of key features of demand without the large support assumption also appears to be new to the literature.

Although the generality of our model makes our results strong, it also brings some limitations.

¹In a fully parametric context, a similar point is suggested by Berry, Levinsohn, and Pakes (2004).

One is a constraint on the types of out-of-sample counterfactuals that can be identified. This is a constraint inherent to nonparametric models: functional form assumptions will often be required if one is to extrapolate outside the support of the data generating process. A second limitation is that our model lacks sufficient structure to permit full characterizations of welfare. In particular, we do not assume a structure that would enable one to track a given consumer’s position in the distribution of indirect utilities across environments. Thus, although we will be able to identify changes for one particular definition of utilitarian social welfare, other welfare measures will often require additional assumptions. We provide additional discussion of these limitations below.

In the following section we set up the model and define the structural features of interest. In section 3 we then demonstrate our basic line of argument for a simple case: binary choice with exogenous characteristics. Our main results are given in section 4, which addresses multinomial choice, allowing endogeneity. We then move to discussion of several important extensions. In section 5 we discuss the case in which the large support condition fails. Section 6 presents testable restrictions of key maintained hypotheses. In sections 7 and 8 we discuss identification using market level data and data from a single market. Having presented our results, we are then able to place our contribution within the context of the large literature on identification of multinomial choice models. After doing this in section 9, we conclude in section 10.

2 Model

2.1 Setup

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

to identification, and our explicit reference to markets provides a way to discuss this clearly.

In applications to demand it is important to model consumers as having the option to purchase none of the products considered (see, e.g., Bresnahan (1981), Anderson, DePalma, and Thisse (1992), Berry (1994) and Berry, Levinsohn, and Pakes (1995)). We represent this by choice $j = 0$ and assume $0 \in \mathcal{J}_t \forall t$. Choice 0 is often referred to as the “outside good.” We denote the number of “inside goods” by $J_t = |\mathcal{J}_t| - 1$.² Each inside good j has observable (to us) characteristics x_{jt} , which may include price. Prices, of course, will generally be correlated with product-specific unobservables. Unobserved choice characteristics are characterized by an index ξ_{jt} , which may also vary across markets. We will assume that ξ_{jt} has an atomless marginal distribution in the population.

Each consumer i in market t is associated with a vector of observables z_{ijt} . The j subscript on z_{ijt} allows the possibility that some characteristics are consumer-choice specific—e.g., interactions between consumer demographics and product characteristics (say, family size and automobile size) or other consumer-specific choice characteristics (say, driving distance to retailer j from consumer i ’s home). For most of our results we will require at least one such measure for each $j \geq 1$, although in section 7 we explore the case in which only market level data are available.

We consider a random utility model. Consumers face no uncertainty themselves, but from the perspective of an outsider the preferences of any individual are viewed as random (e.g., Luce (1959), Block and Marschak (1960), McFadden (1974), Manski (1977)), with the usual interpretation that this reflects unobserved consumer-specific tastes for products and/or characteristics. Fix a market t and the values of $\{(x_{jt}, \xi_{jt}, z_{ijt})\}_{j \in \mathcal{J}_t}$. Let $(\Psi, \mathcal{F}, \mathbb{P})$ denote a probability space.³ Each consumer i ’s preferences are then assumed to be described by conditional indirect utilities

$$v_{ijt} = u(x_{jt}, \xi_{jt}, z_{ijt}, \psi_{it}) \quad \forall i, j \in \mathcal{J}_t \tag{1}$$

²In applications with no “outside choice” our approach can be adapted by normalizing preferences and characteristics relative to those for a given choice. The same adjustment applies when characteristics of the outside good vary across markets in observable ways—something we abstract from for simplicity.

³We depart from the standard notation Ω for the sample space here due to conflicting standard notations and our own inability to distinguish between w and ω .

where $\psi_{it} \in \Psi$, and u is a measurable function.

Implicit in this formulation is a standard restriction that the random variation in the conditional indirect utilities is i.i.d. across individuals and markets. We mark this restriction explicitly with the following.

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

The invariance of \mathbb{P} to i describes the sampling structure: conditional on $\{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}$, unobservable variation in the preferences of different individuals reflects independent draws of the elementary outcome ψ_{it} from Ψ . This does not rule out all within market correlation in consumers' preferences conditional on the observables $\{x_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}$, since ξ_{jt} can be interpreted as an unobserved market-level taste for good j .

The invariance to \mathcal{J}_t and $\{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t}$ (and thus to t) reflects the standard view of preferences as stable, rather than varying with the choice set. In particular, the utility from consuming one good does not vary with the set of goods offered in the market but not consumed.

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

Example 1 *A model within our general framework is the standard linear random coefficients model*

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

where $\beta_{it} = \beta(\psi_{it})$, $\epsilon_{ijt} = \epsilon_{it}I_j$, $\epsilon_{it} = (\epsilon_1(\psi_{it}), \dots, \epsilon_J(\psi_{it}))$, and I_j is a J -vector with j th component equal to 1 and zeros in all other components. In this case, Assumption 1 allows an arbitrary joint distribution of $(\beta_{it}, \epsilon_{it})$ but requires that $(\beta_{it}, \epsilon_{it})$ have the same joint distribution for all i , t , and $\{(x_{jt}, \xi_{jt}, z_{ijt})\}_{j=1 \dots J}$. In this example, it is common to think of ϵ_{ij} as a “product specific taste”, which is modeled as a random coefficient on an x variable (product characteristic) that is defined as a product-specific dummy. In our set up, the ϵ 's can be freely

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

correlated across j , but we require (for example) that the marginal distribution of ϵ_{i1} does not vary as other elements of the choice set change.⁵

Each consumer i maximizes her utility by choosing good j whenever

$$u(x_{jt}, \xi_{jt}, z_{ijt}, \psi_{it}) > u(x_{kt}, \xi_{kt}, z_{ikt}, \psi_{it}) \quad \forall k \in \mathcal{J}_t - \{j\}. \quad (2)$$

Denote consumer i 's choice by

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

Let $z_{ijt} = (z_{ijt}^{(1)}, z_{ijt}^{(2)})$, with $z_{ijt}^{(1)} \in \mathbb{R}$. Let $\mathbf{z}_{it}^{(1)}$ denote the vector $(z_{i1t}^{(1)}, \dots, z_{iJ_t t}^{(1)})'$ and $\mathbf{z}_{it}^{(2)}$ the matrix $(z_{i1t}^{(2)}, \dots, z_{iJ_t t}^{(2)})'$. We will require that for each possible $\mathbf{z}_{it}^{(2)}$, there exist a representation of preferences with the form

$$\tilde{u}_{ijt} = \phi_{it} z_{ijt}^{(1)} + \tilde{\mu}(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \psi_{it}) \quad \forall i, j = 1, \dots, \mathcal{J}_t \quad (3)$$

for some function $\tilde{\mu}$ that is strictly increasing and continuous in ξ_{jt} , and with the random coefficient $\phi_{it} = \phi(\psi_{it}) > 0$.⁶ Here we have imposed two restrictions: (i) additive separability in a “vertical” component, $z_{ijt}^{(1)}$, of z_{ijt} , (ii) monotonicity in ξ_{jt} . We show in section 6 that both restrictions have testable implications.

We rely on the separability restriction to provide a mapping between units of (latent) utility and units of (observable) choice probabilities.⁷ Because unobservables have no natural order, monotonicity in ξ_{jt} would be without loss of generality if consumers had homogeneous tastes for characteristics, as in standard multinomial logit, nested logit, and multinomial probit models. With heterogeneous tastes for choice characteristics, monotonicity imposes a restriction that ξ_{jt}

⁵In many applications, the number of products will vary across markets. In this case, we can think of the outcome ψ as generating all the ϵ 's that would be necessary in the market with the largest possible number of products.

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

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

be a “vertical” rather than “horizontal” choice characteristic. Thus, all consumers agree that (all else equal) larger values of ξ_{jt} are preferred. Of course, our specification does allow heterogeneity in tastes for ξ_{jt} , just as this is permitted for the vertical characteristic $z_{ijt}^{(1)}$. Furthermore, we allow a different representation (3) for each value of $\mathbf{z}_{it}^{(2)}$.⁸

We need to make several normalizations in order to obtain a unique representation of preferences from (3). First, because unobservables have no natural units we may normalize the location and scale of ξ_{jt} and assume without loss that it has a uniform marginal distribution on $(0, 1)$. We must also normalize the location and scale of utilities. Without loss, we normalize the scale of consumer i 's (ordinal) utility using his marginal utility from $z_{ijt}^{(1)}$, yielding the representation

$$u_{ijt} = z_{ijt}^{(1)} + \frac{\tilde{\mu}(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \psi_{it})}{\phi_{it}} \quad \forall i, j = 1, \dots, J^t.$$

Letting

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

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

$$u_{ijt} = z_{ijt}^{(1)} + \mu(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \psi_{it}) \quad \forall i, j = 1, \dots, J^t. \quad (4)$$

To normalize the location we set $u_{i0t} = 0 \forall i, t$. Treating the utility from the outside good as non-stochastic is without loss of generality here, since choices in (2) are determined by differences in utilities and we have not restricted correlation in the random components of utility across choices.

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

⁸Athey and Imbens (2007) point out that the assumption of a scalar vertical unobservable ξ_{jt} can lead to testable restrictions in some models. In our model, if there were no variation across j in $z_{ijt}^{(1)}$ holding consumer characteristics fixed, consumers with the same $\mathbf{z}_{it}^{(2)}$ but different $\mathbf{z}_{it}^{(1)}$ must rank (probabalistically) any products with identical observable characteristics the same way, as they point out. Their observation does not apply to our model in general, however. For example, conditional indirect utilities of the form $v_{ijt} = \xi_{jt} + z_{ijt}^{(1)}\beta_{it}$ are permitted by our model and do not lead to their testable restriction. We discuss testable restrictions of our more general model in section 6.

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

$$\begin{aligned} u_{ijt} &= x_{jt}\beta_{it} + \xi_{jt} + \epsilon_{ijt} \\ \beta_{it}^k &= \beta_0^k + \beta_u^{k0}\nu_{it}^k + \sum_r z_{it}^r \beta_o^{kr} \quad k = 1, \dots, K \end{aligned}$$

where x_{jt} are auto characteristics, z_{it}^r are consumer characteristics, ϵ_{ijt} is assumed distributed type 1 extreme value, each ν_{it}^k is a standard normal deviate, and all stochastic components are i.i.d. Here β_0^k, β_u^{k0} , and β_o^{kr} are all parameters of our function μ in (4).

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

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

with ϵ_{ijt} distributed type I extreme value.

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

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

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

2.2 Observables and Structural Features of Interest

When we discuss the case of endogenous choice characteristics we will require excluded instruments, which we denote by \tilde{w}_{jt} . The observables consist of $(y_{it}, \{x_{jt}, \tilde{w}_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t})_{i,t}$. To discuss

identification, we treat their joint distribution as known.⁹

The observables directly reveal the conditional choice probabilities

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

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

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

For many economic questions motivating demand estimation, less information is required. For example, to discuss cross-price elasticities, equilibrium pricing or market shares under counterfactual ownership or cost structures, knowledge of the demand structure itself is adequate. Identification of demand naturally requires less from the model and/or data than identification of the underlying distribution of preferences. In the multinomial choice setting, demand is fully characterized by the *structural choice probabilities*

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

These conditional probabilities are not directly observable from (5) because of the unobservables ξ_{jt} , which are typically correlated with at least some elements of x_{jt} (e.g., price). Some of our results below address identification of these choice probabilities.

⁹Loosely speaking, we consider the case of a large number of markets, each with a large number of consumers.

2.3 Some Limitations of the Model

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

A second issue more special to the demand application concerns welfare. Our specification of preferences has sufficient structure to characterize changes in the distribution of utilities in meaningful units and, therefore, of utilitarian social welfare. In particular, (4) incorporates quasilinearity preferences.¹¹ However, (4) lacks the structure required for welfare analysis that depends on the *distribution* of welfare changes. This is because we identify the distribution of utility, but do not attempt to separately identify the functional form of μ and the distribution of ψ . Identification of Pareto improvements, for example, will require additional restrictions enabling one to link an individual consumer's position in the distribution of utilities before a policy change to that after. An example of a model with such structure is the standard linear random coefficients model with independent choice-specific taste shocks.

It should not be surprising that some welfare calculations require stronger assumptions than those necessary to identify demand, and our results will help to clarify which questions require these additional restrictions and which do not. Nonetheless, this will be an important limitation in some applications, and we are currently exploring identification of models with sufficient

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

¹¹In the case of binary choice, this characteristic could be income, in which case our utility functions will be money-metric. Our results will then imply identification of changes in aggregate compensating/equivalent variation in income units. If income enters preferences through the function μ in (4), 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.

structure to address such welfare questions. Since these models typically impose *more* structure on the problem than our current assumptions, progress on this front seems fairly likely.

3 Binary Choice with Exogenous Characteristics

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

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

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

We consider identification under the following assumptions.

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

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

Assumption 2 states that we consider the case of exogenous observables. This assumption is relaxed in the following section. A “large support” condition like Assumption 3 is standard in the econometrics literature on nonparametric and semiparametric identification of discrete choice models (e.g., Manski (1985), Matzkin (1992), Matzkin (1993), Lewbel (2000)).¹³ We relax this assumption in section 5. Here we show that Assumptions 1-3 are sufficient for identification.

¹²In the case of demand estimation with endogenous prices, identification arguments using control variates do not appear to be applicable in general. This is because in most models the endogenous $x_{jt}^{(1)}$ (price) is chosen by a firm that has observed all the cost and demand “shocks” in the model, not just its own demand shock ξ_{jt} . This violates the usual requirement that the endogenous right-hand-side variable be one-to-one with a scalar unobservable, conditional on observables (see, e.g., Imbens and Newey (2006)). An exception is the case of binary choice with no cost shocks. For binary response models, Blundell and Powell (2004) consider identification and estimation of a linear semiparametric model using a control function approach.

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

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

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

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

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

we then see that Assumption 3 guarantees that the distribution of $\mu_{it}|t$ (i.e., of μ_{it} in market t) is identified from variation in $z_{it}^{(1)}$ within market t . Denote this distribution by $F_{\mu_{it}|t}(\cdot)$. This argument can be repeated for all markets t .

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

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

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

$$\delta_t = \Delta(x_t, \xi_t) \tag{8}$$

for some function Δ that is strictly increasing in its second argument. To show identification of Δ , for $\tau \in (0, 1)$ let $\delta^\tau(x_t)$ denote the τ th quantile of $\delta_t|x_t$ across markets. By the strict monotonicity of Δ in ξ_t , this quantile is unique and, by (8) and the normalization of ξ_t

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

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

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

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

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

Theorem 1 *Consider the binary choice setting with preferences given by (4). Under Assumptions 1–3, the distribution of u_{it} conditional on any $(\mathcal{J}_t, x_t, \xi_t, z_{it})$ in their support is identified.*

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

4 Main Results

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

Assumption 4 $\text{supp } \mathbf{z}_{it}^{(1)} | \mathbf{x}_t, \mathbf{z}_{it}^{(2)} = \mathbb{R}^{J_t} \forall \mathbf{x}_t, \mathbf{z}_{it}^{(2)}$.

Without the assumption $(x_{1t}, \dots, x_{Jt}) \perp (\xi_{1t}, \dots, \xi_{Jt})$, we will need instruments. To state our instrumental variables assumptions, let $x_{jt} = \left(x_{jt}^{(1)}, x_{jt}^{(2)} \right)$, where $x_{jt}^{(1)}$ denotes the endogenous characteristics. We then let $w_{jt} = \left(x_{jt}^{(2)}, \tilde{w}_{jt} \right)$ denote the vector of instrumental variables.

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

Assumption 6 *Conditional on (x_{jt}, w_{jt}) , ξ_{jt} has a continuous bounded pdf $f_\xi(\cdot | x_{jt}, w_{jt})$. For any bounded function $\Delta(x_{jt})$, let $\mathcal{T}(x_{jt}, w_{jt}) = \int_0^1 f_\xi(\tau \Delta(x_{jt}) | x_{jt}, w_{jt}) d\tau$. Assume (i) ... ***additional technical conditions to be added*** ... and (ii) if $E[\Delta(x_{jt}) \mathcal{T}(x_{jt}, w_{jt}) | w_{jt}] = 0$ a.s., then $\Delta(x_{jt}) = 0$ a.s.*

Assumption 5 provides the necessary excludability condition on the instruments \tilde{w}_{jt} . Assumption 6 provides the remaining IV condition, ensuring that the instruments induce sufficient variation in the endogenous variables. This is a particular type of “bounded completeness” condition, which we take from Chernozhukov and Hansen (2005) (Appendix C).¹⁴ With these assumptions, we can prove the following result

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

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

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

¹⁴They discuss sufficient conditions. We also consider an alternative to Assumption 6 below.

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

$$\begin{aligned}\delta_{jt} &= E[\mu_i(x_{jt}, \xi_{jt}) | t] \\ &= \Delta(x_{jt}, \xi_{jt})\end{aligned}\tag{9}$$

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

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

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

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

Instead of (3), suppose each consumer i 's conditional indirect utilities can be represented by

$$\tilde{u}_{ijt} = \beta_{it} z_{ijt}^{(1)} + \tilde{\mu}(x_{jt}, z_{ijt}^{(2)}, \psi_{it}) + \gamma_{it} \xi_{jt}, \quad j = 1, \dots, \mathcal{J}_t.\tag{11}$$

This imposes a restriction relative to (3) but is still more general than standard specifications of utility in the literature (e.g., Lewbel (2000)). A representation of preferences equivalent to

(11) is then given by

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

where now

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

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

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

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

and scale normalization

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

Although (14) may appear unusual, it is without loss of generality. It defines units of the unobservable ξ_{jt} by fixing the mean marginal rate of substitution between $z_{ijt}^{(1)}$ and ξ_{jt} .

To prove identification of the joint distribution of $\{u_{ijt}\}_j$ conditional on $\{x_{jt}, z_{ijt}, \xi_{jt}\}_j$, first note that the argument in the proof of Theorem 2 remains valid through equation (9). Recall that we have fixed the value of $\left(z_{i1t}^{(2)}, \dots, z_{iJt}^{(2)} \right)$ and dropped these arguments. With the separable structure (13) and the normalization (14) now we have

$$\delta_{jt} = E [\mu (x_{jt}, \xi_{jt}, \psi_{it}) | t] = \Delta (x_{jt}) + \xi_{jt} \quad (15)$$

for some function Δ . It is straightforward to confirm that, under Assumption 5, the following “completeness” condition is equivalent to identification of the function Δ (Newey and Powell (2003)) from observation of $(\delta_{jt}, x_{jt}, \tilde{w}_{jt})$.

Assumption 7 For all functions $\Delta (x_{jt})$ with finite expectation, $E [\Delta (x_{jt}) | w_{jt}] = 0$ implies $\Delta (x_{jt}) = 0$ a.s.

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

Theorem 3 *Under the utility representation (12), suppose Assumptions 1, 4, and 7 hold. Then the joint distribution of $\{u_{ijt}\}_{j \in \mathcal{J}_t}$ conditional on any $\left(\mathcal{J}_t, \{(x_{jt}, z_{ijt}, \xi_{jt})\}_{j \in \mathcal{J}_t}\right)$ 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 Δ and therefore of each ξ_{jt} . The remainder of the proof then follows that of Theorem 2 exactly, beginning with (10). ■

The completeness condition (Assumption 7) is the analog of the rank condition in linear models. It requires that variation in \tilde{w}_{ijt} induce sufficient variation in $x_{jt}^{(1)}$ to reveal $\Delta(x_{jt})$ at all points x_{jt} . Lehman and Romano (2005) give standard sufficient conditions. Severini and Tripathi (2006) point out that this condition is equivalent to the following: for any bounded function $f(x_{jt})$ such that $E[f(x_{jt})] = 0$ and $\text{var}(f(x_{jt})) > 0$, there exists a function $h(\cdot)$ such that $f(x_{jt})$ and $h(\tilde{w}_{jt})$ are correlated. Additional intuition can be gained from the discrete case: as shown by Newey and Powell (2003), when x_{jt} and w_{jt} have discrete support $(\hat{x}^1, \dots, \hat{x}^K) \times (\hat{w}^1, \dots, \hat{w}^L)$, completeness corresponds to a full rank condition on the matrix $\{\sigma_{kl}\}$ where $\sigma_{kl} = \Pr(x_{jt} = \hat{x}^k | w_{jt} = \hat{w}^l)$.

5 Limited Support

The large support assumption (Assumption 4) in the preceding section is standard and demonstrates that sufficient variation in the vector $(z_{i1}^{(1)}, \dots, z_{ij_t}^{(1)})$ 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 will be needed: if the observable data can move choice probabilities only through a subset of the simplex, the most we can hope for is to identify the joint distribution of utilities on a subset of *their* support.

Of course, one would like to know what can be learned from more limited variation in the data. We explore this question here and show that much more limited variation can be sufficient to identify the structural choice probabilities $\rho_j(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t})$ at all points of

support. As discussed above, these choice probabilities are sufficient by themselves for many questions that motivate estimation of discrete choice demand models. Given the observability of p_{ijt} , the essential step is demonstrating identifiability of the choice-specific unobservables ξ_{jt} . In addition, we show a type of continuity: there is a natural sense in which moving from our limited support condition to the full support condition moves the identified features of the model smoothly toward the full identification results of the preceding section.

We first explore identifiability of the structural choice probabilities without the separability assumption in (4) and without any support assumption whatsoever. Thus far we have obtained positive results only for the binary choice case in this situation. This offers one motivation for exploration of conditions that rely on the separable structure used above but much weaker support requirements. In particular, instead of requiring $z_{it}^{(1)}$ to move choice probabilities through the entire simplex in each market, we require only that there be *some* choice probability vector that is attainable in every market by the appropriate choices of the vectors $(z_{i1t}^{(1)}, \dots, z_{iJt}^{(1)})$.

5.1 Identification of ξ_{jt} With No Support Condition

Consider the general specification of preferences in (1). In the case of binary choice the consumer selects the inside good if

$$u(x_t, \xi_t, z_{it}, \psi_{it}) > 0. \tag{16}$$

Note that we have dropped the earlier requirement of additive separability in $z_{it}^{(1)}$. In fact, there need not exist any individual-choice specific observables at all. Under Assumption 1, the probability of the event (16) can be written

$$\pi_{it} = \rho(x_t, z_{it}, \xi_t) \tag{17}$$

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

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

A key to this result can be seen from the case without endogeneity. For that case, consider the distribution $F_\pi(\cdot|x_t, z_{it})$ of π_{it} across markets conditional on (x_t, z_{it}) . The market with choice probability at the τ th quantile of the distribution of $F_\pi(\cdot|x_t, z_{it})$ is the market with $\xi_t = \tau$ (recalling that ξ_t is $u(0, 1)$). Thus, one identifies each ξ_t by inverting equation (17):

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

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

$$\pi_{ijt} = \rho_j(\mathcal{J}_t, x_{1t}, \dots, x_{Jt}, z_{i1t}, \dots, z_{iJt}, \xi_{1t}, \dots, \xi_{Jt}) \quad j = 1 \dots J. \quad (18)$$

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

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

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

5.2.1 Binary Choice

As before, we will begin with the binary choice case to illustrate our insights most simply. Here we will discuss two results. We first consider the general specification of preferences in (4) above. We then consider the a more restrictive specification (analogous to (12)) that appears to be more useful for the multinomial case. For both results we make the following “common

choice probability” assumption:¹⁵

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

This is a much weaker requirement than the full support assumption, and is likely to be satisfied in many applications.

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

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

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

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

Then each z_t^τ is the τ th quantile of the random variable $-\mu_{it} = -\mu(x_t, \xi_t, \psi_{it})$ conditional on t — i.e., on (x_t, ξ_t) . Thus each z_t^τ is a function only of (x_t, ξ_t) :

$$z_t^\tau = \zeta(x_t, \xi_t; \tau) \tag{19}$$

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

Identification of the function $\zeta(\cdot; \tau)$ and, therefore, of each ξ_t , then follows from (19) as in the preceding sections, again using the nonparametric instrumental variables result of Chernozhukov and Hansen (2005). With each ξ_t known, the observable choice probabilities $p(w_t, z_{it})$ reveal the structural choice probabilities

$$\rho(x_t, \xi_t, z_{it}) = \Pr(y_{it} = 1 | x_t, \xi_t, z_{it}) \tag{20}$$

¹⁵Implicitly we require here a continuous (region of) support for $\mu(x_t, \xi_t, z_{it}^{(2)}, \psi_{it}) | x_t, \xi_t, z_{it}^{(2)}$ to guarantee uniqueness.

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

Theorem 5 *In the binary choice model with preferences given by (4), suppose Assumptions 1, 5, 6, and 8 hold. Then the structural choice probabilities $\rho(x_t, \xi_t, z_{it})$ are identified at all points (x_t, ξ_t, z_{it}) in their support.*

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

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

in the limit—i.e., with sufficient variation in $z_{it}^{(1)}$ to make every $\tau \in (0, 1)$ a common choice probability—we are back to full identification as in Theorem 2.

Additive ξ As before, we can replace Assumption 6 with Assumption 7, and for this we require linear separability of ξ_t , as in (15). Here we also impose a further restriction on the utility specification in (11)—in particular, that $\frac{\gamma_{it}}{\beta_{it}}$ to equal one *for each consumer* rather than merely *in expectation*. After normalizing by γ_{it} , this leads to the representation

$$u_{ijt} = z_{ijt}^{(1)} + \mu(x_{jt}, z_{ijt}^{(2)}, \psi_{it}) + \xi_{jt} \quad \forall i, j = 1, \dots, \mathcal{J}_t. \quad (21)$$

This involves a significant restriction on preferences relative to our previous results, although a similar restriction has been required for the most general prior results for identification of linear semiparametric models with heteroskedasticity and endogeneity (e.g., Lewbel (2000)). Thus, our most restrictive specification (21) is still a generalization relative to the literature even without our relaxation of the the large support assumption here.

Theorem 6 *In the binary choice model with preferences given by (21), suppose Assumptions 1, 5, 7, and 8 hold. Then the structural choice probabilities $\rho(x_t, \xi_t, z_{it})$ are identified at all points (x_t, ξ_t, z_{it}) in their support.*

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

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

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

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

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

$$z_t^\tau = \zeta(x_t; \tau) - \xi_t. \tag{22}$$

Equation (22) can then be used as before to identify $\zeta(\cdot; \tau)$ (and therefore ξ_t) using the results of Newey and Powell (2003). As in the preceding section, this is sufficient to determine the structural choice probabilities $\rho(x, \xi_t, z_{it}) = \Pr(y_{it} = 1 \mid x_t, \xi_t, z_{it})$ at all points (x_t, ξ_t, z_{it}) of support.

■

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

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

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

5.3 Multinomial Choice

For the multinomial case we generalize the previous common choice probability assumption as follows:

Assumption 9 For some $q = (q_0, q_1, \dots, q_J) \in \nabla^J$, there exists for each market t a vector $\mathbf{z}_t^q = (z_{1t}^q, \dots, z_{Jt}^q) \in \text{supp}(z_{i1t}^{(1)}, \dots, z_{iJt}^{(1)})$ such that for all j

$$q_j = \Pr(y_{it} = j \mid x_{1t}, \dots, x_{Jt}, z_{i1t}, \dots, z_{iJt})_{\mathbf{z}_{it}^{(1)} = \mathbf{z}_t^q}.$$

Here we will maintain the more restrictive representation of preferences in (21), where

$$u_{ijt} = z_{ijt}^{(1)} + \mu(x_{jt}, z_{ijt}^{(2)}, \psi_{it}) + \xi_{jt} \quad \forall i, j = 1, \dots, \mathcal{J}_t. \quad (23)$$

With this specification, choice probabilities depend only on the sums

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

rather than on each $z_{ijt}^{(1)}$ and ξ_{jt} separately.

Theorem 7 In the multinomial choice model with preferences given by (23), suppose Assumptions 1, 5, 7, and 9 hold. Then the structural choice probabilities $\rho_j(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t})$ are identified at all $(\mathcal{J}_t, \{x_{jt}, \xi_{jt}, z_{ijt}\}_{j \in \mathcal{J}_t})$ in their support.

Proof. Fixing the vector $\mathbf{z}_t^{(2)}$,

$$\begin{aligned} p_{ijt} &= \Pr(y_{it} = j \mid x_{1t}, \dots, x_{Jt}, \xi_{1t}, \dots, \xi_{Jt}, z_{i1t}, \dots, z_{iJt}) \\ &= \Pr(y_{it} = j \mid x_{1t}, \dots, x_{Jt}, \lambda_{i1t}, \dots, \lambda_{iJt}) \\ &= \Pr(\mu(x_{jt}, \psi_{it}) + \lambda_{ijt} \geq \mu(x_{kt}, \psi_{it}) + \lambda_{ikt} \quad \forall k). \end{aligned}$$

Each p_{ijt} is strictly increasing and continuous in λ_{ijt} . So we know from Berry (1994) and/or Berry and Pakes (2007) that for any vector of choice probabilities $q = (q_1, \dots, q_{J_t})$ and any \mathbf{x}_t

vector there is a unique vector $\lambda(\mathbf{x}_t, q) = (\lambda_1(\mathbf{x}_t, q), \dots, \lambda_{J_t}(\mathbf{x}_t, q))$ such that

$$q_j = \Pr(y_{it} = j | \mathbf{x}_t, \lambda(\mathbf{x}_t, q)) \quad \forall j.$$

From the definition of $\lambda(\mathbf{x}_t, q)$, we know that at the common choice probability vector q and the associated values of \mathbf{z}_{jt}^q defined in Assumption 9

$$z_{jt}^q + \xi_{jt} = \lambda_j(\mathbf{x}_t, q) \quad \forall j$$

i.e.,

$$z_{jt}^q = \lambda_j(\mathbf{x}_t, q) - \xi_{jt} \quad \forall j. \tag{24}$$

This equation can then be used as before to identify the function $\lambda_j(\cdot, q)$ and each ξ_{jt} using the results for the additively separable nonparametric IV regression in Newey and Powell (2003). As demonstrated above, knowledge of all ξ_{jt} identifies the structural choice probability functions.

■

6 Testable Restrictions

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

Our assumption that preferences can be represented by conditional indirect utilities in which $z_{ijt}^{(1)}$ enters in an additively separable fashion (with positive coefficient) has two implications on the reduced form choice probabilities.

Theorem 8 *If preferences are characterized by (4), then under Assumption 1 and,*

- (i) $\Pr(y_{it} = j | \mathcal{J}_t, \{x_{jt}, \tilde{w}_{jt}, z_{ikt}\}_{k \in \mathcal{J}_t})$ is strictly increasing in $z_{ijt}^{(1)}$; and
- (ii) $\partial p_{ijt} / \partial z_{ijt}^{(1)} = - \sum_{k \neq j} \partial p_{ikt} / \partial z_{ikt}^{(1)}$.

Proof. Both results are immediate from the equation

$$p_{ijt} = \Pr \left(z_{ijt}^{(1)} + \mu \left(x_{jt}, \xi_{jt}, z_{ijt}^{(2)}, \psi_{it} \right) > z_{ikt}^{(1)} + \mu \left(x_{kt}, \xi_{jt}, z_{ijt}^{(2)}, \psi_{it} \right) \quad \forall k \neq j \right)$$

since, under Assumption 1, variation in $\left\{ z_{jt}^{i(1)} \right\}_{j \in \mathcal{J}_t}$ does not change the joint distribution of $\left\{ \mu \left(x_{kt}, \xi_{jt}, z_{ijt}^{(2)}, \psi_{it} \right) \right\}_{j \in \mathcal{J}_t}$. ■

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

Theorem 9 *In the binary choice model with preferences given by (4), suppose Assumptions 1, 5, 6, and 8 hold. Then $\xi_t(z_t^\tau; \tau, x_t)$ must be strictly decreasing in z_t^τ across markets.*

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

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

Example 5 *Suppose $\mu(x_t, \xi_t, \phi_{it}) = -\nu_{it}\xi_t$, with $\nu_{it} \sim N(0, 1)$. Take $\tau > 1/2$ and consider the set of markets in which $\xi_t(z_t^\tau; \tau, x_t) > 0$.¹⁶ Recall that each z_t^τ is observable and defined such that $\Pr(\nu_{it}\xi_t < z_t^\tau) = \tau$. Letting Φ denote the standard normal CDF, this requires*

$$\Phi \left(\frac{z_t^\tau}{\xi_t} \right) = \tau \quad \forall t. \tag{25}$$

Therefore, by construction, $\frac{z_t^\tau}{\xi_t}$ will take the same value in every market. Since each z_t^τ must also be positive, this requires a strictly positive correspondence between z_t^τ and ξ_t across markets. Thus, $\xi_t(z_t^\tau; \tau, x_t)$ will violate the testable restriction of the Theorem 9.

Theorem 10 *In the binary choice model with preferences given by (4), suppose Assumptions 1, 5, 6, and 8 hold. In addition, suppose that for distinct τ and τ' in the interval $(0, 1)$, for every*

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

market t there exists a unique $z_t^\tau \in \text{supp } z_{it}^{(1)}$ such that $\Pr(y_{it} = 1 | z_{it}^{(1)} = z_t^\tau) = \tau$ and a unique $z_t^{\tau'} \in \text{supp } z_{it}^{(1)}$ such that $\Pr(y_{it} = 1 | z_{it}^{(1)} = z_t^{\tau'}) = \tau'$. Then $\xi_t(z_t^\tau; \tau, x_t) = \xi_t(z_t^{\tau'}; \tau', x_t)$ for all t .

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

Example 6 *forthcoming*

7 Identification with Market Level Data

In many applications one is forced to work without micro data linking choices to individual characteristics. To consider identification in such environments, suppose preferences are characterized by conditional indirect utility functions

$$\tilde{u}_{ijt} = \tilde{u}(x_{jt}, \xi_{jt}, \psi_{it}) \quad j \in \mathcal{J}_t \quad (26)$$

and that the only observables are \mathcal{J}_t , x_{jt} , and the market shares p_{jt} of each good j in market t .

We begin by pointing out that the discussion in section 5.1 applies directly in this case: there we made no use of the micro data z_{ijt} in the identification argument and, as noted there, the argument is the same when there are no z_{ijt} at all. Thus, Theorem 4 shows that in the case of binary choice, ξ_t and, therefore, the structural choice probabilities are identified without any support requirements whatsoever. As before, we conjecture that a similar result may be obtainable in the multinomial choice environment.

To explore additional results, let $x_{jt} = (x_{jt}^{(1)}, x_{jt}^{(2)})$ and suppose that preferences can be represented by conditional indirect utilities of the form

$$u_{ijt} = x_{jt}^{(1)} + \mu(x_{jt}^{(2)}, \xi_{jt}, \psi_{it}). \quad (27)$$

Assume that the set of markets can be partitioned into *market groups* Γ such that for all $t \in \Gamma$, $(x_{jt}^{(2)}, \xi_{jt}) = (x_{j\Gamma}^{(2)}, \xi_{j\Gamma})$. One natural example of such an environment is that of a national

industry (e.g., the U.S. automobile industry) in which 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), etc. To show identification here we make three assumptions similar to those used in section 4.

Assumption 10 $\xi_{jt} \perp (w_{jt}, x_{ijt}^{(1)}) \forall j, t$.

Assumption 11 *Conditional on $(x_{jt}^{(2)}, w_{jt})$, ξ_{jt} has a continuous bounded pdf $f_\xi(\cdot | x_{jt}^{(2)}, w_{jt})$. For any bounded function $\Delta(x_{jt}^{(2)})$, let $\mathcal{T}(x_{jt}^{(2)}, w_{jt}) = \int_0^1 f_\xi(\tau \Delta(x_{jt}^{(2)}) | x_{jt}^{(2)}, w_{jt}) d\tau$. Assume (i) ... additional technical conditions to be added ... and (ii) if $E[\Delta(x_{jt}^{(2)}) \mathcal{T}(x_{jt}^{(2)}, w_{jt}) | w_{jt}] = 0$ a.s., then $\Delta(x_{jt}^{(2)}) = 0$ a.s.*

Assumption 12 *supp $(x_{1t}^{(1)}, \dots, x_{J_t t}^{(1)}) | (x_{1t}^{(2)}, \dots, x_{J_t t}^{(2)}) = \mathbb{R}^{J_t} \forall t$.*

Assumption 12 is different from the parallel Assumption 4 in requiring sufficient variation in a product characteristic rather than an individual-product observable. The role of this assumption is the same, however: to trace out the distribution of the random component of (27) within each market group.

Under these assumptions, variation in $x_{jt}^{(1)}$ across market groups at the limit $x_{j't}^{(1)} \rightarrow -\infty \forall j' \neq j$ identifies the distribution of $\mu_i(x_{j\Gamma}^{(2)}, \xi_{j\Gamma})$ exactly as in the previous sections. Letting $\delta(x_{j\Gamma}^{(2)}, \xi_{j\Gamma}) = E[\mu_i(x_{j\Gamma}^{(2)}, \xi_{j\Gamma}) | x_{j\Gamma}^{(2)}, \xi_{j\Gamma}]$, identification of the function $\delta(x_{j\Gamma}^{(2)}, \xi_{j\Gamma})$ (and therefore each $\xi_{j\Gamma}$) follows exactly as in the previous sections. With each $\xi_{j\Gamma}$ and the distribution of $\mu_i(x_{j\Gamma}^{(2)}, \xi_{j\Gamma})$ known, the joint distribution of $\{u_{ijt}\}_{j \in \mathcal{J}_t}$ is uniquely determined at any $(\mathcal{J}_t, \{(x_{jt}, \xi_{jt})\}_{j \in \mathcal{J}_t})$ in their support..

Theorem 11 *Consider the multinomial choice model with market level data and preferences given by (27). If Assumptions 1, 10, 11 and 12 are satisfied, then the joint distribution of $\{u_{ijt}\}_{j \in \mathcal{J}_t}$ conditional on any $(\mathcal{J}_t, \{(x_{jt}, \xi_{jt})\}_{j \in \mathcal{J}_t})$ in their support.*

8 Identification in a Single Market

forthcoming

9 Relation to the Literature

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

Identification of linear random-coefficients binary choice models has been considered by Ichimura and Thompson (1998) and Gautier and Kitamura (2007). Both consider models with mutually independent taste shocks (the random coefficients and the consumer-choice specific shock). Briesch, Chintagunta, and Matzkin (2005) consider multinomial choice, allowing some generalizations of the linear random coefficients model but maintaining mutual independence of all taste shocks. All of these impose additional assumptions as well. None relaxes our requirement of linearity in at least one characteristic. None allows choice-specific unobservables or endogenous choice characteristics.

Recent work considering endogenous choice characteristics includes Lewbel (2000), Honoré and Lewbel (2002), Hong and Tamer (2004), Blundell and Powell (2004), Lewbel (2005), and Magnac and Maurin (2007). These all consider linear semiparametric models with a single additively separable taste shock for each choice. All but the two papers by Lewbel limit attention to binary response models, which can be significantly simpler and, for example, can be amenable to control function methods in the context of demand estimation.

An important distinction between our work and much of the prior literature is our neglect of estimation. Although our constructive identification proofs may suggest new nonparametric or semiparametric estimation approaches, additional work is clearly needed. Moving from identification to estimation typically requires imposition of appropriate smoothness conditions at a minimum. In practice additional restrictions may need to be imposed on our model to obtain estimators with desirable properties.

10 Conclusion

We have examined the identifiability of multinomial choice models of demand. Our framework allows a nonparametric distribution-free specification of random utility with endogenous choice characteristics, unknown heteroskedasticity, and correlation between taste shocks. Our results are obtained using standard assumptions from the literature, and we have shown that for some questions the standard “large support” assumption can in fact be relaxed considerably while preserving identification of the structural choice probabilities that determine the answers to many questions that motivate estimation of demand in differentiated products markets.

One reason we have been able to make this progress is our direct focus on the joint distributions of conditional indirect utilities. Given the characteristics $((x_{1t}, z_{i1t}, \xi_{1t}), \dots, (x_{J_t t}, z_{iJ_t t}, \xi_{J_t t}))$ of any choice set, this is a joint distribution of J_t random variables. Contrast this with a linear random coefficients model without unobserved product characteristics,

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

In that case the structure consists of the joint distribution of $\{\beta_i, \epsilon_{ij}\}_j$, which has $J_t + \dim x$ components. Given that the observable (conditional on $((x_{1t}, z_{i1t}, \xi_{1t}), \dots, (x_{J_t t}, z_{iJ_t t}, \xi_{J_t t}))$) is the J_t -dimensional vector $(p_{i1t}, \dots, p_{iJ_t t})$, it is perhaps not surprising that additional assumptions must be imposed to obtain identification. Considering the nonparametric joint distribution of utilities directly turns out to provide sufficient natural limits the dimensionality of the identification problem without substantial restrictions on preferences themselves. Our progress under limited support arises from the observation that for many purposes the structural choice probabilities alone determine the answers to the economic questions of interest. Of course, as we have mentioned already, our focus on the primitives of a very general model also implies some limits on the types of questions that can be answered. One contribution of our work is therefore to help focus the question of what kinds of assumptions enable identification of the welfare measures that are not determined by the primitives of our model.

References

- ANDERSON, S., A. DEPALMA, AND F. THISSE (1992): *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge MA.
- ATHEY, S., AND P. A. HAILE (2002): “Identification of Standard Auction Models,” *Econometrica*, 70(6), 2107–2140.
- ATHEY, S., AND G. W. IMBENS (2007): “Discrete Choice Models with Multiple Unobserved Choice Characteristics,” Discussion paper, Harvard.
- BAYER, P., F. FERREIRA, AND R. MCMILLAN (2007): “A Unified Framework for Measuring Preferences for Schools and Neighborhoods,” *Journal of Political Economy**Review of Economic Studies*, 115(5), 588–638.
- BERRY, S. (1994): “Estimating Discrete Choice Models of Product Differentiation,” *RAND Journal of Economics*, 23(2), 242–262.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 60(4), 889–917.
- (2004): “Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market,” *Journal of Political Economy*, 112(1), 68–105.
- BERRY, S. T., AND A. PAKES (2007): “The Pure Characteristics Demand Model,” Discussion paper, Yale University.
- BLOCK, H., AND J. MARSCHAK (1960): “Random Orderings and Stochastic Theories of Responses,” in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. by I. Olkin, S. Ghurye, W. Hoeffding, W. G. Mado, and H. B. Mann. Stanford University Press.
- BLUNDELL, R. W., AND J. L. POWELL (2004): “Endogeneity in Semiparametric Binary Response Models,” *Review of Economic Studies*, 71, 655–679.

- BRESNAHAN, T. (1981): “Departures from Marginal Cost Pricing in the American Automobile Industry,” *Journal of Econometrics*, 17, 201–227.
- BRIESCH, R. A., P. K. CHINTAGUNTA, AND R. L. MATZKIN (2005): “Nonparametric Discrete Choice Models with Unobserved Heterogeneity,” Discussion paper, Northwestern University.
- CAPPS, C., D. DRANOVE, AND M. SATTERTHWAITE (2003): “Competition and Market Power in Option Demand Markets,” *RAND Journal of Economics*, 34(5), 737–763.
- CHERNOZHUKOV, V., AND C. HANSEN (2005): “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73(1), 245–261.
- DOMENICH, T., AND D. MCFADDEN (1975): *Urban Travel Demand: A Behavioral Analysis*. North Holland, Amsterdam.
- GAUTIER, E., AND Y. KITAMURA (2007): “Nonparametric Estimation in Random Coefficients Binary Choice Models,” Discussion paper, Yale.
- GOLDBERG, P. K. (1995): “Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry,” *Econometrica*, 63(4), 891–951.
- GUADAGNI, P. M., AND J. D. C. LITTLE (1983): “A Logit Model of Brand Choice Calibrated on Scanner Data,” *Marketing Science*, 2(3), 203–238.
- HASTINGS, J., D. STAIGER, AND T. KANE (2007): “Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery,” Discussion paper, Yale University.
- HECKMAN, J. J., AND B. E. HONORÉ (1990): “The Empirical Content of the Roy Model,” *Econometrica*, 58, 1121–1149.
- HO, K. (2007): “Insurer-Provider Networks in the Medical Care Market,” Discussion paper, Columbia University.
- HONG, H., AND E. TAMER (2004): “Endogenous Binary Choice Mode with Median Restrictions,” *Economics Letters*, pp. 219–224.

- HONORÉ, B. E., AND A. LEWBEL (2002): “Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors,” *emet*, 70(5), 2053–2063.
- ICHIMURA, H., AND T. S. THOMPSON (1998): “Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution,” *Journal of Econometrics*, 86(2), 269–95.
- IMBENS, G., AND W. NEWEY (2006): “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” Discussion paper, M.I.T.
- KENNAN, J., AND J. WALKER (2006): “The Effect of Expected Income on Individual Migration Decisions,” Discussion paper, University of Wisconsin-Madison.
- LEHMAN, E., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*. Springer, New York, 3 edn.
- LEWBEL, A. (2000): “Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables,” *Journal of Econometrics*, 97, 145–177.
- (2005): “Simple Endogenous Binary Choice and Selection Panel Model Estimators,” Discussion paper, Boston College.
- LUCE, R. D. (1959): *Individual Choice Behavior*. Wiley.
- MAGNAC, T., AND E. MAURIN (2007): “Identification and Information in Monotone Binary Models,” *Journal of Econometrics*, 139, 76–104.
- MANSKI, C. F. (1977): “The Structure of Random Utility Models,” *Theory and Decision*, 8, 229–254.
- (1985): “Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator,” *Journal of Econometrics*, 27, 313–333.
- (1988): “Identification of Binary Response Models,” *Journal of the American Statistical Association*, 83(403), 729–738.

- MATZKIN, R. (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.
- McFADDEN, D. (1974): “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Frontiers of Econometrics*, ed. by P. Zarembka. Academic Press, New York.
- NEWKEY, W. K., AND J. L. POWELL (2003): “Instrumental Variable Estimation in Nonparametric Models,” *Econometrica*, 71(5), 1565–1578.
- RIVERS, D. (1988): “Heterogeneity in Models of Electoral Choice,” *American Journal of Political Science*, 32(3), 737–757.
- SEVERINI, T. A., AND G. TRIPATHI (2006): “Some Identification Issues in Nonparametric Models with Endogenous Regressors,” *Econometric Theory*.