Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models

Jeremy T. Fox
University of Michigan and NBER

Amit Gandhi
University of Wisconsin-Madison*

August 2013

Abstract

Multinomial choice and other nonlinear models are often used to estimate demand. We show how to nonparametrically identify the distribution of unobservables, such as random coefficients, that characterizes the heterogeneity among consumers in multinomial models. In particular, we provide general identification conditions for a class of nonlinear models and then verify these conditions using the primitives of the multinomial choice model. We require that the distribution of unobservables lie in the class of all distributions with finite support, which under our most general assumptions is an infinite dimensional space. We allow prices to be endogenous and also study the case where consumers purchase multiple products at once, with bundle-specific prices. We show how identification leads to the consistency of a nonparametric estimator.

*Thanks to Stephane Bonhomme, Steven Durlauf, James Heckman, Salvador Navarro, Philip Reny, Susanne Schennach, Azeem Shaikh, Christopher Taber, Harald Uhlig and Edward Vytlacil for helpful comments. Also thanks to seminar participants at many workshops. Fox thanks the National Science Foundation, the Olin Foundation, and the Stigler Center for financial support. Thanks to Chenchuan Li for research assistance. Our email addresses are jtfox@umich.edu and agandhi@ssc.wisc.edu.

1 Introduction

Heterogeneity among decision makers, be they firms or consumers, is a critical feature of economic life. A classic example is demand for differentiated products. We observe many competing products, each with different characteristics, being offered for sale. The presence of these differentiated products suggest that consumers have heterogeneous preferences for product characteristics. In order to model the demand for these products, one approach is to estimate the distribution of consumers' heterogeneous preferences.

In applications in empirical industrial organization, the demand for differentiated products is typically modeled using a multinomial choice model (McFadden, 1973). In multinomial choice, a consumer picks one of a finite number of alternatives, each with measured product characteristics including price. In modern empirical practice, the key unknown object to researchers is a distribution of unobserved heterogeneity (Domencich and McFadden, 1975; Heckman and Willis, 1977; Hausman and Wise, 1978; Boyd and Mellman, 1980; Cardell and Dunbar, 1980). This distribution of unobservables captures the heterogeneous preferences consumers have for the product characteristics. Sometimes these heterogeneous preferences are called random coefficients. We will use the term "heterogenous unobservables" to encompass random coefficients and other forms of heterogeneity. In the highest generality, each realization of the heterogenous unobservable can index a different realization of a utility function over product characteristics, so that different consumers have different utility functions.

This paper presents tools for identifying and estimating the distributions of heterogeneous unobservables in a class of economic models that includes the multinomial choice demand model with random coefficients. Indeed, this is the first paper to show that such a distribution of random coefficients entering linear indices in utilities is identified in a multinomial choice model without logit errors. Also, we provide machinery to identify the distribution of heterogeneous unobservables in a class of nonlinear models and then use the machinery to identify the distribution of heterogeneous unobservables in the multinomial choice model.

We focus on the common data scheme in applied microeconomics where the researcher has access to cross sectional rather than panel data. In the multinomial choice environment, we observe different consumers facing different menus of observed product characteristics. We characterize each consumer in the population by a heterogeneous behavioral parameter $\theta \in \Theta$, the heterogenous unobservable. The heterogeneous unobservable θ has distribution G. We seek the nonparametric identification and estimation of G. In a high level of generality, θ can be a utility function of observed product characteristics.

The nonparametric identification and estimation of distributions of unobservables is well-understood in the case of the linear regression model with random coefficients. Let $y = a + x'b$, where x is a real vector with continuous support and y is a real outcome variable. In the linear random coefficient model, $\theta = (a, b)$ is a heterogeneous vector of intercept and slope parameters. One example is a Cobb-Douglas production function in logs, where y is log output, x is a vector of log inputs, a is total factor productivity, and b is the vector of input elasticities. Compared to a model without random coefficients, the random coefficient regression model allows the effect of changing inputs x to vary across firms: b is a heterogeneous parameter. Some firms in the same industry may have labor-intensive technologies and others may have capital-intensive technologies. In the random coefficient production function model, the object of interest is the distribution of random coefficients $G(a, b)$, or the joint distribution of total factor productivities and input elasticities. Knowledge of $G(a, b)$ tells us the distribution of production functions in an industry, which is an argument to answering many policy questions, for example the effects of taxing some particular input. Beran and Millar (1994) and Beran (1995) first demonstrated the nonparametric identification of the distribution of the slope and intercept parameters in the linear model. Hoderlein, Klemelä and Mammen (2010) focus mainly on estimation but show the nonparametric identification of $G(a, b)$ in the linear model while allowing endogeneity through an auxiliary, linear instrumental variables equation without random coefficients.

Two papers that extend identification of the distribution of unobservables to a nonlinear setting with continuous outcomes are Liu (1996) and our own, more general work in Fox and Gandhi (2011). Fox and Gandhi study a nonlinear model where $y = g(x, \theta)$, where $g(\cdot, \cdot)$ is a known, real analytic function of x for each θ, θ is a finite or an infinite dimensional unobservable, y is a real outcome variable and x is a vector with support on the reals or rationals. For example, $g(\cdot, \cdot)$ might represent some nonlinear production function, such as the constant elasticity of substitution (CES). It might also represent an aggregate demand curve. The object of interest is $G(\theta)$, the distribution of the heterogeneous unobservables. Fox and Gandhi allow endogeneity through a pricing equation that may have its own infinite dimensional heterogeneous unobservable. This pricing equation is shown to be consistent with standard models of price setting behavior, like Bertrand-Nash oligopoly.

The results in Fox and Gandhi apply to the case of modeling aggregate market shares or quantities. However, they are less useful for estimating models of consumer choice using micro data. Often in applied practice, models of consumer choice involve consumers making discrete choices. For example, a consumer may be modeled as choosing one out of a finite number of competing brands. In a multinomial choice setting, the relationship between y and x is discontinuous. Indeed, continuous changes in x and θ can lead to discrete changes in the choice made by the same consumer. There are existing results for discrete choice models with two choices. Ichimura and Thompson (1998) and Gautier and Kitamura (2013) study identification and estimation two-outcome choice models where random coefficients enter linear indices. Gautier and Kitamura allow for endogeneity using an auxiliary equation without random coefficients.

Our result for the multinomial choice model nests the case of multiple purchases with bundle-specific prices: a consumer can choose two or more of the discrete alternatives and may have preferences over the bundles themselves. We finally explore identification of the distribution of random coefficients in the multinomial choice model where regressors such as price can be endogenous. Our solution to endogeneity uses an auxiliary pricing function (with heterogeneous unobservables) and hence parallels our solution to endogeneity for the case of continuous outcomes in Fox and Gandhi (2011).

Our primitive assumptions help clarify the role of data in showing identification in the multinomial choice model. Specifically, we rely on a product specific regressor that enters additively into the utility for a product and that can take on arbitrarily large values. Further, the sign of the random coefficient on this particular regressor is common across consumers. The additivity, common sign and large support ensure that all consumers, indexed by heterogeneous unobservables θ, will purchase any particular product at some value of the regressors in the data. Indeed, our property for generic economic models includes the idea that different consumers, or heterogeneous unobservables θ, must take different actions at some regressors x in order for the distribution of θ, G, to be identified. We do not require these assumptions on the other product characteristics and instruments in the model. Other papers making use of so-called "special regressor" arguments include Ichimura and Thompson (1998), Lewbel (1998), Lewbel (2000), Matzkin (2007) and Berry and Haile (2010). We assume full independence between regressors or instruments and the heterogeneous unobservables, instead of Lewbel's assumption of only mean independence. As pointed out by Magnac and Maurin (2007) and Khan and Tamer (2010), identification of homogeneous parameters (which we do not study) under only mean independence is sensitive to learning the tails of the distribution of unobservables, as means are sensitive to the exact masses in the tails. Fox, Kim, Ryan and Bajari (2012) do not require special regressors, but require the parametric assumption that additive errors for each discrete choice are i.i.d. with the type I extreme value distribution, which leads to the random coefficients logit model.[1]

Note that results on special regressors can be used to identify the distribution of utility value realizations for all choices in a multinomial choice model, as in Berry and Haile (2010). Our results our only useful relative to the literature when the researcher is interested in learning the joint distribution of random coefficients in a linear index model, a joint distribution of finite dimensional heterogenous parameters entering a known utility function, or the joint distribution of utility functions for all choices. We describe the difference with say Berry and Haile (2010) in the objects of identification in more detail below.

Our identification results for the multinomial choice model rely on other possibly strong assumptions, which are also maintained to a lesser degree in Fox and Gandhi (2011). Most importantly, we require that the true distribution of unobservables takes unknown but finite support in the, likely uncountable or even infinite dimensional, space of possible unobservables. Thus, we learn the number of support points and the location of the support points in identification. Because the space where the unobservable lives can be infinite dimensional, the space of distributions of unobservables is itself

[1]Chiappori and Komunjer (2009) discuss some assumptions under which they can show the identification of a multinomial choice model without additive regressors. Manski (2007) considers identification of counterfactual choice functions and structural preferences when there are a fixed number of decision problems and no observed regressors x. Hoderlein (2009) considers a binary choice model with endogenous regressors. Briesch, Chintagunta and Matzkin (2010) study a model with a scalar unobservable that enters nonseparably into the utility of each choice.

allowed to be infinite dimensional. Clearly it would be nice to relax the unknown but finite support assumption, but we believe our results to be on the frontier of known results for the multinomial choice model.

Second, we require that each distinct unobservable utility function over observable product characteristics take on different values in an open set of observed product characteristics. In other words, every two realizations of the heterogeneous unobservables have different utility values for some product characteristics within an arbitrarily small set of product characteristics. Each realized utility function, if the utility function is the heterogeneous unobservable, being real analytic in observed product characteristics is sufficient for this property. We are less concerned with the restrictiveness of this assumption as analytic functions nest utility functions that are linear in the regressors and the random coefficients, a common specification.

Our identification results show that only the true distribution of random coefficients could generate a dataset with an infinite number of observations. We also discuss nonparametric estimation of the distribution of random coefficients. Here we turn to the computationally simple, nonparametric estimator of Fox, Kim, Ryan and Bajari (2011) and Fox, Kim and Yang (2013). To make the discussion of how identification leads to a consistent estimator self-contained, we cite a consistency theorem from Fox et al. (2013) and show how the identification results in this paper show that this estimator is consistent.

Section 2 previews our identification results for a simple multinomial choice model where random coefficients enter into linear indices, a common empirical specification. Section 3 reviews notation for a generic economic model, states the property for identification and proves, by contradiction, that the property is sufficient for identification of the distribution of random coefficients in the generic model. Section 4 applies the framework for the generic economic model to our case of primary interest, identifying the distribution of heterogeneous unobservables in the multinomial choice model, which nests the linear random coefficients in Section 2. Section 5 extends our results for multinomial choice to allow for price endogeneity. Section 6 explores estimation and shows how to use our identification arguments with a particular nonparametric estimator.

Mathematically we demonstrate the identification of the distribution of unobservables in three multinomial choice models: those in Sections 2, 4, and 5, with the model of Section 2 again being a common empirical specification of the more general model in Section 4. For brief illustration, we also consider two trivial examples of models in Section 3.1. All three identification results for multinomial choice models and two identification results for trivial examples are applications of the identification machinery for a generic, not necessarily multinomial choice, model in Section 3. Therefore, we need to show how the notation for the generic model maps into the five separate models. As a consequence, notation is used differently for each of the five specific models as, for example, the key notion of the heterogeneous unobservable differs across the models.

2 Random Coefficients Entering Linear Indices

This section introduces a simple multinomial choice model where random coefficients enter linear indices and previews our results on identification for this example.

Let there be J inside goods and choice 0, the outside option of no purchase. Each good could be a competing product. Let $\mathcal{Y}$ be the set of J inside goods plus the outside good 0. Let x_j be a vector of observable (in the data) characteristics for product j. In our main results, these will be characteristics with continuous support. Characteristics will vary across choice situations (say geographic markets), or if product characteristics are interacted with consumer demographics, consumers. We suppress the consumer and market subscripts for simplicity. Let $x = (x_1, \ldots, x_J)$ collect the characteristics for all choices.

The utility to choice j is $x_j'\beta + \epsilon_j$, where ϵ_j is an additive intercept in j's utility reflecting the horizontal taste for a product and β is a vector of random coefficients. Both ϵ_j and β are specific to the consumer, so that consumers have heterogeneous marginal valuations of the observed product characteristics in x_j We do not assume that each ϵ_j has mean 0, so ϵ_j subsumes the role of a homogeneous (non-random) intercept for product j. Each element of the vector of random coefficients β gives the consumer's marginal utility for the corresponding product characteristic in x_j. The outside good indexed by 0 has its utility normalized to 0. The vector $\theta = (\beta, \epsilon_1, \ldots, \epsilon_J)$ collects all the unobservable terms for a consumer and is the key heterogeneous unobservable whose joint distribution we will identify. Given this, the discrete choice $y \in \mathcal{Y}$ is given by

$$y = f(x, \theta) \equiv \arg\max_{j \in \mathcal{Y}} \left\{ \left\{ x_j'\beta + \epsilon_j \right\}_{j=1}^{J}, 0 \right\},$$

where 0 corresponds to the normalized utility of choice 0 and $f(x, \theta)$ picks the choice with the highest utility for a consumer with heterogeneous unobservables θ and a menu of observed product characteristics x.

For each consumer, the researcher observes (x, y), or the menu of observed characteristics and the discrete choice. Using these data on the choices of different consumers, the goal is to identify the true distribution $G^0(\theta)$, or how the random coefficients β and intercepts ϵ_j in the vector θ vary across consumers. Given this structure, the probability of a consumer picking choice j given observed product characteristics x is

$$\Pr_{G^0}(j \mid x) = \int 1\left[j = f(x, \theta)\right] dG^0(\theta),$$

or the integral of the purchase decision over the distribution of consumer additive errors and random coefficients. This can also be seen as a market share equation. This choice probability imposes that the heterogeneous unobservables θ are independent of the product characteristics x. We delay a discussion of the endogeneity of product characteristics (like price) until Section 5.

We set $x_j'\beta + \epsilon_j = v_j'\beta + r_j + \epsilon_j$ for $x_j = (v_j, r_j)$, where v_j is a vector of characteristics and r_j is a special, scalar characteristic called a choice-specific special regressor. In our use, the term special regressor will refer to a regressor with a 1) common sign across consumers and 2) support equal to $\mathbb{R}$, the real line. The fact that the coefficient of r_j is ± 1 is a scale normalization: the units of utility are not identifiable so we express them in units of r_j for each consumer.[2] The assumptions that the sign of r_j is common across consumers and that the support is large are restrictive. The binary and multinomial choice literature usually uses such a special regressor to gain point identification (Manski, 1988; Ichimura and Thompson, 1998; Lewbel, 1998, 2000; Briesch et al., 2010; Gautier and Kitamura, 2013; Matzkin, 2007; Berry and Haile, 2010). The intuition, which we will make explicit below, comes from the need that different consumer heterogeneous unobservables θ must take different actions y at different choice sets x. Say $J = 1$ and ϵ_1 is very high, but different, for two consumers. Then both consumers will buy the inside good for typical values of v_1 and r_1. Only by making r_1 very small can we ensure that one or both of the consumers switches to buy the outside good of 0, when $v_1'\beta + r_1 + \epsilon_1 < 0$. So we need a regressor r_j that can shift the payoff of each choice to be arbitrarily large or small, as the utility of each choice from the additive component ϵ_j can also be arbitrarily large or small.

What results on identification exist in the literature using this special regressor assumption? Ichimura and Thompson (1998) study the case of one inside good or $J = 1$, and show that $G^0(\theta)$ is identified using the Cramér and Wold (1936) theorem. Gautier and Kitamura (2013) provide a computationally simpler estimator for the same model. These two papers have not been generalized to multinomial choice. In several papers, Matzkin studies identification in discrete choice models where there are limits to the correlation of utility across choices so that our example model is not a special case (Briesch et al., 2010; Matzkin, 2007). Lewbel (2000) and, in a more general context circulated contemporaneously with the first draft of our paper, Berry and Haile (2010) show that the distribution of $\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_J)$ where

$$\tilde{u}_j = v_j'\beta + \epsilon_j$$

is identified conditional on $v = (v_1, \ldots, v_J)$, the non-special regressors. This means that the distribution $H(\tilde{u} \mid v)$ of the vector of the non-r *utility values* $\tilde{u}$ conditional on v is identified from variation in the special regressors $r = (r_1, \ldots, r_J)$. The argument for this is simple. The choice probability for the outside good 0 can be written

$$\begin{aligned}\Pr(0 \mid x) = \Pr(0 \mid v, r) &= \int 1\left[v_j'\beta + r_j + \epsilon_j \leq 0 \,\forall\, j = 1, \ldots, J\right] dG^0(\beta, \epsilon_1, \ldots, \epsilon_J) \\ &= \Pr(\tilde{u}_1 \leq -r_1, \ldots, \tilde{u}_J \leq -r_J \mid v, r) = H(-r \mid v).\end{aligned}$$

[2]The sign of r_j can be easily identified. If r_j enters utility with a positive sign, the observed conditional (on product characteristics) probability of picking choice j will increase when r_j increases.

The latter object is the cumulative distribution function (CDF) of $\tilde{u}$, $H(\tilde{u} \mid v)$, evaluated at the point $\tilde{u} = -r$. Thus, if the vector of special regressors r varies flexibly over $\mathbb{R}^J$, we can trace out the CDF of the J utility values $\tilde{u}$, conditional on v. Knowing the distribution of utility values is enough to predict market shares in sample, i.e. for values of v and r observed in the data.

Our paper attempts to get at the distribution $G^0(\theta) = G^0(\beta, \epsilon_1, \ldots, \epsilon_J)$ of random coefficients and random intercepts. This is the primitive of the multinomial choice model in this section. Identifying the true primitive is useful in several contexts. For example, identifying $G^0(\theta)$ lets one predict the market shares for values of the product characteristics in v not observed in the estimation sample. This is the new goods problem: we identify $\Pr_{G^0}(j \mid v^\star, r)$ for menus of product characteristics $v^\star$ not in the original data's support of $\Pr_{G^0}(j \mid v, r)$. Also, identifying $G^0(\theta)$ lets one find the distribution of utility differences between two choice situations. We identify the joint distribution of $\left(v_j^{\star\prime}\beta + \epsilon_j + r_j^\star - \left(v_j'\beta + \epsilon_j + r_j\right)\right)_{j=1}^J$, the utility improvement for each of the J products if the product characteristics in $x = (v, r)$ change to $x^\star = (v^\star, r^\star)$. We can also identify the distribution of

$$\max_{j \in \mathcal{Y}} \left\{ \left\{ x_j^{\star\prime}\beta + \epsilon_j \right\}_{j=1}^J, 0 \right\} - \max_{j \in \mathcal{Y}} \left\{ \left\{ x_j'\beta + \epsilon_j \right\}_{j=1}^J, 0 \right\}, \tag{1}$$

the differences in maximized utility values, one version of a "treatment effect" for changing x to $x^\star$.

By contrast, the distribution of utility values $H(\tilde{u} \mid v)$ does not assign utility to particular realizations of heterogeneous unobservables θ, and so a researcher cannot calculate (1). The lack of a distribution of heterogeneous unobservables and consequently utility functions prevents the researcher from computing a distribution of welfare changes, a major use of structural demand models. Of course, the distribution of utility values is sufficient to identify mean utility differences, as means are linear operators and so calculating them requires only the distribution of utility values at both x and $x^\star$:

$$E\left[\left(v_j^{\star\prime}\beta + \epsilon_j + r_j^\star\right) - \left(v_j'\beta + \epsilon_j + r_j\right)\right] = E\left[v_j^{\star\prime}\beta + \epsilon_j + r_j^\star\right] - E\left[v_j'\beta + \epsilon_j + r_j\right].$$

Again, previous results allow this computation only for v and $v^\star$ in the original support of v.

The main result of our paper, for this example, is that $G^0(\theta)$ is identified from data on (x, y), with r having full support. As we are identifying a more primitive object than Lewbel (2000) and Berry and Haile (2010), we impose a stronger assumption on the support of θ. In particular, we impose that θ has unknown finite support. In other words, for this example we assume that G^0 is known to lie in the space $\mathcal{G}$ of distributions that admit a finite support on $\mathbb{R}^{\dim(\theta)}$. We learn the number of and location of the heterogeneous unobservables in the support of G^0 as part of identification. This assumption does not nest the random coefficients logit, as the type I extreme value distribution on the ϵ_j's has continuous support and the β often is assumed to have a normal distribution, which also has continuous support. However, we do not impose any parametric assumptions, such as the type I extreme value and normal distributions. We also do not impose that β is independent of the additive errors $\epsilon_1, \ldots, \epsilon_J$ or that the additive errors are i.i.d. across choices.

3 A Generic Economic Model

We now consider a generic economic model that can be described by a tuple $(\Theta, \mathcal{X}, \mathcal{Y}, f, \mathcal{G})$. The space Θ of heterogeneous unobservables θ represents, in a demand model, the feasible set of consumer preferences admitted by the model. Each heterogeneous unobservable θ can index an infinite-dimensional object such as a function; in a later section each consumer will have its own, heterogenous utility function. The fact that Θ can be infinite-dimensional is why the title of the paper refers to "nonparametric".

The set $\mathcal{X}$ denotes the set of economic environments in the support of the data generating process. The set $\mathcal{Y}$ is the (measurable) outcome space. The known function $f : \mathcal{X} \times \Theta \to \mathcal{Y}$ maps a consumer's heterogeneous unobservable $\theta \in \Theta$ and economic environment $x \in \mathcal{X}$ to an outcome $y = f(x, \theta) \in \mathcal{Y}$. The joint distribution of outcomes and environments (y, x) is identified from the i.i.d., cross sectional data. What remains to be identified is the distribution of heterogeneous unobservables $G \in \mathcal{G}$ in the population, where $\mathcal{G}$ is a set of probability measures over Θ. Not surprisingly, proving the identification of G will require assumptions on all aspects of the model $(\Theta, \mathcal{X}, \mathcal{Y}, f, \mathcal{G})$.

As discussed previously, we let $\mathcal{G}$ be the class of distributions with finite support on Θ. Therefore, we can represent each $G \in \mathcal{G}$ by the notation

$$\left(N, (\theta_n, w_{\theta n})_{n=1}^{N}\right),$$

where the number of heterogeneous unobservables N is an unknown to be identified, as are the identities of each support point θ_n and the weights on each support point w_{θ_n}. Of course $\sum_{n=1}^{N} w_{\theta_n} = 1$. Because Θ can be a infinite-dimensional space, $\mathcal{G}$ constitutes an infinite dimensional space of distributions. Indeed, even one support point θ_n of a distribution in $\mathcal{G}$ is infinite dimensional when Θ is infinite dimensional.

Let $A \subseteq \mathcal{Y}$ be a measurable subset of the outcome space. Assuming statistical independence (for now) between the heterogeneous unobservable θ and the covariates x, if $G^0 \in \mathcal{G}$ is the true distribution of heterogeneous unobservables in the population, we have that

$$\Pr_{G^0}(A \mid x) = G^0(\{\theta \in \Theta \mid f(x, \theta) \in A\}) = \sum_{n=1}^{N^0} w_{\theta_n}^0 1\left[f(x, \theta_n^0) \in A\right]. \tag{2}$$

Thus the distribution G^0 is identified up to the measure it assigns to sets of the form $I_{A,x} = \{\theta \in \Theta \mid f(x, \theta) \in A\}$, which are indexed by a point x and a set $A \subseteq \mathcal{Y}$. The problem is whether the class of such sets $I_{A,x}$ is rich enough to point identify G^0 within the class of distributions $\mathcal{G}$.

To state the identification problem precisely, let $\Pr(\cdot \mid x)$ be a probability measure over $\mathcal{Y}$ for a given value $x \in \mathcal{X}$ of the environment. Let $P = \{\Pr(\cdot \mid x) \mid x \in \mathcal{X}\}$ denote a collection of such probability measures over all possible economic environments and let $\mathcal{P}$ denote the set of all such collections P. Then we can view (2) as a mapping $L : \mathcal{G} \to \mathcal{P}$. We will say the distribution G^0 is *identified* if L is a one-to-one map. That is, for any $G, G' \in \mathcal{G}$ and $G \neq G'$, there exists an experiment in the data (A, x) where $A \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$ such that $\Pr_G(A \mid x) \neq \Pr_{G'}(A \mid x)$, where $\Pr_G(\cdot \mid x)$ and $\Pr_{G'}(\cdot \mid x)$ are the images of G and G' respectively under L.

Identification requires showing that, for any two potential distribution of heterogeneous unobservables, there always exists an experiment in the data (A, x) that can empirically distinguish between these distributions. For specific models, the researcher in some sense needs flexible supports for $\mathcal{Y}$ and $\mathcal{X}$ to find such an experiment (A, x) to distinguish any two distributions. In the multinomial choice model, $\mathcal{Y}$ is a list of exclusive, discrete choices. Not surprisingly then, identification of G^0 in the multinomial choice model will require flexibility in $\mathcal{X}$, the support of the product characteristics.

Definition 1. For any set of heterogeneous unobservables $T \subset \Theta$, and for any $A \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$, the ***I*-set** $I^T_{A,x}$ is defined as

$$I^T_{A,x} \equiv \{\theta \in T \mid f(x, \theta) \in A\}.$$

An I-set is the set of heterogeneous unobservables within an arbitrary subset of heterogeneous unobservables $T \subset \Theta$ whose response is in the set A at the covariates x. Thus, I-sets are indexed in part by the observables in the data: the dependent variable set A and the independent regressor vector x. I-sets are strictly a property of the underlying economic choice model and are independent of the particular distribution of unobserved heterogeneity G. The set of heterogeneous unobservables T should not be seen as the support of the true distribution, but merely an arbitrary set of heterogeneous unobservables. Further, the full set of feasible heterogeneous unobservables Θ within the model can be quite distinct from the subset T considered in the definition of an I-set.

We now state and prove an identification result for this generic economic model.

Theorem 1. *Let a generic model $(\Theta, \mathcal{X}, \mathcal{Y}, f, \mathcal{G})$ satisfy the following property.*

- *For any finite set $T \subset \Theta$, there exists a pair (A, x) such that corresponding I-set $I^T_{A,x}$ is a singleton.*

Then the true $G^0 \in \mathcal{G}$ is identified.

The theorem states that the researcher can identify the number N, identity $\{\theta_n\}_{n=1}^N$ and the mass $\{w_n\}_{n=1}^N$ of the support points. The property in this theorem is useful because we can prove that it holds using primitive assumptions in empirical models of some interest, such as the multinomial choice model. The property in the theorem also clarifies the role of dependent and independent variables in identification. Their role is to find situations x where different heterogeneous unobservables θ take different actions y.

Proof. Recall that identification requires showing that the mapping $L : \mathcal{G} \to \mathcal{P}$ defined by (2) is one to one. Thus for $G^0, G^1 \in \mathcal{G}$ with $G^0 \neq G^1$, we must have that $\Pr_{G^0}(A \mid x) \neq \Pr_{G^1}(A \mid x)$ for some $A \subseteq \mathcal{Y}, x \in \mathcal{X}$. In particular, for any $P \in L(\mathcal{G})$, we show that $L\left(G^0\right) = L\left(G^1\right) = P$ implies $G^0 = G^1$.

We can represent any $G \in \mathcal{G}$ by a pair (T, p), where $T = \{\theta_1, \ldots, \theta_N\} \subset \Theta$ is a finite set of heterogeneous unobservables and the probability vector $w = \{w_\theta\}_{\theta \in T}$ comprises non-negative masses that sum to one over T. Given the representation (T, w) for $G \in \mathcal{G}$, we can express (2) as

$$\Pr_G(A \mid x) = \sum_{\theta \in I^T_{A,x}} w_\theta. \tag{3}$$

If G^0 is represented by (T^0, w^0) and G^1 is represented by (T^1, w^1), then we can redefine w^0 and w^1 so that G^0 and G^1 are represented by (T, w^0) and (T, w^1) respectively, where $T = T^0 \cup T^1$ (for example, if $\theta \in T - T^0$, then set $w^0_\theta = 0$). T is still finite. Moreover if we define the vector $\{\pi_\theta\}_{\theta \in T}$ such that $\forall \theta \in T, \pi_\theta = w^0_\theta - w^1_\theta$, then $G^0 = G^1$ if and only if $\pi_\theta = 0$ for all $\theta \in T$.

Our goal is to show that $L\left(G^0\right) = L\left(G^1\right)$ implies $G^0 = G^1$. Observe that $L\left(G^0\right) = L\left(G^1\right)$ implies that for all $A \subseteq \mathcal{Y}$ and $x \in \mathcal{X}$, $\Pr_{G^0}(A \mid x) = \Pr_{G^1}(A \mid x) = \Pr(A \mid x)$, which by (3) implies that

$$\sum_{\theta \in I^T_{A,x}} \pi_\theta = 0, \tag{4}$$

for all I-sets $I^T_{A,x}$. We now show that (4) implies $\pi_\theta = 0$ for all $\theta \in T$. Assume to the contrary that $T_2 = \{\theta \in T \mid \pi_\theta \neq 0\}$ is non-empty. By the property in the statement of the theorem, we can produce a singleton $I^{T_2}_{A,x} = \{\theta^\star\}$. Furthermore, we can re-write (4) as

$$\sum_{\theta \in I^T_{A,x}} \pi_\theta = \sum_{\theta \in I^{T_2}_{A,x}} \pi_\theta + \sum_{\theta \in I^{T-T_2}_{A,x}} \pi_\theta = \sum_{\theta \in I^{T_2}_{A,x}} \pi_\theta = \pi_{\theta^\star} \neq 0,$$

which contradicts (4). Hence it must be that T_2 is empty, and thus $\pi_\theta = 0$ for all $\theta \in T$. □

The above proof is non-constructive.[3] We show how this identification argument can yield a consistent, nonparametric estimator for G in Section 6.

3.1 Trivial Examples

We now present two trivial examples, one of a model that is clearly identified and one of a model that is clearly unidentified. We show that the former model satisfies the property in Theorem 1 and the latter model does not.

[3]The property in Theorem 1 is sufficient for identification, but the property is not necessary. Indeed, our computational experiments suggest that the true distribution of random coefficients can be estimated in at least one model that does not satisfy the property.

Let the model be $y = \theta$, with $\theta \in \Theta \subseteq \mathbb{R}$. The heterogenous unobservable θ is just the dependent variable itself. This distribution of θ in this model is clearly identified because the observed distribution of y is the distribution of θ. This model also satisfies the property in Theorem 1. Let $I_y^T = \{\theta \in T \mid \theta = y\}$. Then choosing any $y^\star = \theta^\star$ for any $\theta^\star \in T$ gives I_y^T equal to just the singleton $\{\theta^\star\}$.

Now consider the model $y = \theta_a + \theta_b$, with $\theta = (\theta_a, \theta_b) \in \mathbb{R}^2$. The joint distribution of θ_a and θ_b in this model is clearly not identified and thus it is instructive to consider where the property in the theorem fails. Consider the I-set $I_y^T = \{\theta \in T \mid \theta_a + \theta_b = y\}$. For a counterexample to the sufficient condition for identification, let $T = \left\{\theta^1, \theta^2\right\}$ such that $\theta_a^1 + \theta_b^1 = \theta_a^2 + \theta_b^2$ but the heterogeneous unobservables are distinct, so $\theta_a^1 \neq \theta_a^2$. For $y^\star = \theta_a^1 + \theta_b^1$, $I_{y^\star}^T = T$ and for any $y \neq y^\star$, $I_y^T = \emptyset$. Therefore, the property in Theorem 1 is not satisfied.

4 Heterogeneous Utility Functions

This section explores multinomial choice in a more general setting than the case of random coefficients entering linear indices in Section 2. The key generalization is that the heterogeneous unobservable θ indexes utility functions over product characteristics. In other words, utility functions are heterogeneous; each consumer has its own, possibly nonlinear utility function for weighting observed product characteristics. We will state a set of assumptions that allow us to apply the identification machinery for a generic model in Theorem 1 to the multinomial choice model.

Consider a consumer θ making a discrete choice from among J products and one outside good. Let $\mathcal{Y} = \{0, 1, \ldots, J\}$, where 0 is the outside good. Each product $j \in \mathcal{Y} - \{0\}$ is characterized by a scalar characteristic $r_j \in \mathbb{R}$. We let $v \in \mathbb{R}^K$ denote the observed characteristics of the consumer and the menu of product characteristics (of the J products) excluding the scalar characteristics in $r = (r_1, \ldots, r_J)$. We call the choice-specific regressors r_j special regressors. We let $x = (v, r) \in \mathbb{R}^{K+J}$ denote the entire menu of consumer and product characteristics, including the scalar characteristics.

We follow the usual special regressor convention that the permissible range of variation in each r_j for $j \in J$ is independent of the product and consumer characteristics v.

Assumption 1. *Let $V \subset \mathbb{R}^K$, the support of v, be a (non-trivial) rectangle. Let $x = (v, r) \in \mathcal{X} = V \times R_1 \times \cdots \times R_J$ where $R_j = \mathbb{R}$ for each $j \in J$.*

A realization of the heterogeneous unobservable $\theta = u = \left(u^1, \ldots, u^J\right)$ is a vector of *functions* of the product characteristics $v \in V$. That is, a realization of the heterogeneous unobservable θ is a vector of functions $u : V \to \mathbb{R}^J$ where each individual function is given by $u^j(v)$. Utility functions are heterogeneous across consumers. The goal is to identify the distribution of the heterogeneous and unobservable utility functions $\theta = u = \left(u^1, \ldots, u^J\right)$.

Assumption 2. *The vector-valued function θ is statistically independent of the observable product characteristics and consumer demographics $x = (v, r)$.*

We discuss endogeneity of the product characteristics in v in Section 5. We need a monotonicity and additive separability assumption for the special regressor r_j.

Assumption 3. *The utility of a heterogeneous unobservable $\theta = u = \left(u^1, \ldots, u^J\right)$ purchasing product j is $u^j(v) + r_j$.*

The utility of the outside good $j = 0$ is normalized to 0. A consumer's response at $x = (v, r)$ is given by the discrete choice that maximizes utility, or

$$y = f(x, \theta) \equiv \arg\max_{j \in \mathcal{Y}} \left\{ \left\{ u^j(v) + r_j \right\}_{j=1}^{J}, 0 \right\}.$$

We restrict attention to utility functions that satisfy a particular no-ties property. Let $\theta(v)$ be the vector of utility values of the inside goods evaluated at v, meaning $\theta(v) = u(v) = \left(u^1(v), \ldots, u^J(v)\right)$. This property simply says that different heterogeneous unobservables have different utility realizations at some point v in an arbitrarily small rectangle V.

Assumption 4. *The heterogeneous unobservable space Θ of feasible utility functions satisfies the following property. For any finite subset of M utility function vectors $\{\theta_1, \ldots, \theta_M\}$, there exists $v \in V$ such that $\theta_c(v) \neq \theta_d(v)$ as vectors for any distinct θ_c and θ_d in $\{\theta_1, \ldots, \theta_M\}$.*

Remark 1. To our knowledge, the only general sufficient condition for Assumption 4 is that each heterogeneous unobservable θ is comprised of J **multivariate real analytic** utility functions u_j. Real analytic functions are equal to a power series in a domain of convergence around each point of evaluation. Appendix A formally defines a real analytic function and proves that Assumption 4 is implied by the space of real analytic functions. We do not formally mention real analytic utility functions in Assumption 4 to emphasize that our identification argument uses no other properties of real analytic functions. Note that the class of real analytic functions includes polynomials and interaction terms. Even if one restricts attention to functional forms where random coefficients enter linear indices as in Section 2, our identification results identify the order of the polynomials and the identity of the interaction terms in the true data generating process. This contrasts our results with Fox, Kim, Ryan and Bajari (2012), who examine the random coefficients logit model and rule out polynomials, including interaction terms.

Example. In the simpler example in Section 2, the utility of choice j is $v_j'\beta + \epsilon_j + r_j$ where $v = (v_1, \ldots, v_J)$. Let $u^j(v) = v_j'\beta + \epsilon_j$. This is a real analytic function in v_j, and so Assumption 4 is implied by this functional form. Mathematically, the case of random coefficients entering linear indices can be seen as a restriction of the space Θ from all real analytic utility (in v) functions to all linear (in v_j) utility functions.

Remark 2. One use of our identification results is to predict choice probabilities and market shares out of sample, meaning for values of product characteristics v not in the original support V in Assumption 1. Specializing equation (2) to this model, the data generating process that we identify is

$$\Pr_G(j \mid x) = \sum_{n=1}^{N} w_{\theta_n} 1 \left[j \in \arg\max_{j \in \mathcal{Y}} \left\{ \left\{ u_n^j(v) + r_j \right\}_{j=1}^{J}, 0 \right\} \right],$$

where $\left(N, (\theta_n, w_{\theta n})_{n=1}^{N}\right)$ is the identified finite-support representation of the true $G(\theta)$ and each support point is $\theta_n = u_n = \left(u_n^1, \ldots, u_n^J\right)$. Therefore, we can evaluate $\Pr_G(j \mid x)$ at any $x = (r, v)$, including $v \notin V$. This ability to do out of sample prediction is not usual for nonparametric methods because it is possible that two different heterogeneous unobservables θ behave identically within V but differently outside of V. In this case, identification analysis within the support V is not useful for learning the behavior of consumers outside of V. In this section, the ability to predict behavior out of the support V arises partly because of Assumption 4: different heterogeneous unobservables $\theta = u$ are required to take different actions within the support V. Otherwise, two heterogeneous unobservables could be observationally equivalent within the support V and there would be no hope of ever identifying their separate frequencies. This paper is about identifying distributions of heterogeneous unobservables, so some assumption along the lines of Assumption 4 will be necessary for this to occur within the support V. Remark 1 argues that a sufficient condition for Assumption 4 is real analytic utility functions. Our results show that a distribution of real analytic utility functions and a distribution of a finite vector of heterogeneous parameters that enter a parametric real analytic utility function can be identified within a small open set V.

Remark 3. Recall that v captures all of the non-r_j product characteristics and the argument to $u^j(v)$ is v, not v_j. In the general model, letting the utility to product j also depend on the characteristics of products $k \neq j$ can capture the idea of context or "menu" effects in consumer choice. Even if such effects are not economically desirable, there is no cost to us in mathematical generality and thus we let the whole menu v enter as an argument to each u^j. The choice-specific scalar r_j, however, enters preferences in an additively separable way (and hence preferences are quasilinear in this scalar characteristic). One example is that r_j could be the price of good j, in which case $u^j(v)$ is the consumer with unobservable heterogeneity θ's reservation price for product j, and utilities are better expressed as $u^j(v) - r_j$. However, r_j could be some non-price product characteristic or, with individual data, an interaction of a consumer and product characteristic, like the geographic distance between a consumer and a store.

Remark 4. Implicit in the quasilinear representation of preferences $u^j(v) + r_j$ is the scale normalization that each realization of unobservable heterogeneity θ's coefficient on r_j is constrained to be 1. The normalization of the coefficient on r_j to be ± 1 is innocuous if each r_j enters utility in the same way for each product; choice rankings are preserved by dividing any heterogeneous unobservable's utilities $u^j(v) + r_j$ by a positive constant. Thus if θ admitted a heterogeneous coefficient $\alpha > 0$, then $\theta_1 = (u, \alpha)$ would have the exact same preferences as $\theta_2 = \left(\frac{u(v)}{\alpha}, 1\right)$. The assumption that r_j has a sign that is the same for each vector of utility functions θ is restrictive. Such a monotonicity restriction on one covariate will be generally needed to show identification of the multinomial choice model using the identification machinery for a generic model in Theorem 1. The sign of r_j could be taken to be negative instead (as in the case where r_j is price), and it is trivial to extend the results to the case where r_j's sign is unknown and constant a priori.

Theorem 2. *Under assumptions 1, 2, 3, and 4, the distribution $G^0 \in \mathcal{G}$ of vectors of utility functions $\theta = u$ in the multinomial choice model is identified.*

Proof. We verify the property in Theorem 1. As in the property, let a finite and arbitrary $T \subset \Theta$ be given, where $T = \{\theta_1, \ldots, \theta_N\}$ and each heterogeneous unobservable $\theta_i = u_i$ is a vector of J utility functions. We consider I-sets, Definition 1, of the form

$$I^T_{0,v,r} = \{\theta \in T \mid f((v, r), \theta) = 0\},$$

which is those heterogeneous unobservables $\theta \in T$ that pick the outside good 0 at $x = (v, r)$. To verify the property in Theorem 1, we will find a $x = (v, r)$ such that $I^T_{0,v,r}$ is a singleton.

Because the set of vectors $\{\theta(v) \mid \theta \in T\}$ is finite, there exists a vector $\underline{\theta}(v)$ that satisfies our definition of a minimal vector in the following footnote.[4] There could be multiple minimal vectors; we focus on one, $\underline{\theta}(v) = \underline{u}(v)$.

Then set the vector $-r = \underline{\theta}(v)$. This means that the vector of product specific utilities

$$\underline{\theta}(v) + r = \left(\underline{u}^1(v) + r_1, \ldots, \underline{u}^J(v) + r_J\right)$$

is equal to the vector of J 0's for the heterogeneous unobservable $\underline{\theta}$. Further, the utility at $x = (v, w)$ is strictly positive for at least one $j > 0$ for all heterogeneous unobservables in T other than $\underline{\theta}$, by our definition of a minimal vector. Now we can lower the vector r by an infinitesimal positive amount so that a consumer with the heterogeneous unobservable $\underline{\theta}$ strictly prefers to purchase the outside good 0 and all other heterogeneous unobservables $\theta_j \in T - \{\underline{\theta}\}$ purchase an inside good at $x = (v, w)$. Thus, $I^T_{0,v,r} = \{\underline{\theta}\}$ is a singleton and we can apply Theorem 1. $\square$

[4] Let S be a set of distinct vectors, each of the form $s = (s^1, \ldots, s^M)$, meaning that each vector s in S has the same number M of (real) elements s^m. We define a minimal vector $\underline{s} \in S$ of the set S to be a vector $\underline{s} \in S$ such that, for each $s \in S - \{\underline{s}\}$, there exists an element index $m(s)$ such that $\underline{s}_{m(s)} < s_{m(s)}$, or that there is some element of the other vector strictly greater than the corresponding element of the minimal vector. It is easy to prove that there exists a minimal vector if S has a finite number of vectors in it. A minimal vector may not be unique. For example, in the set of vectors $\{(1, 0, 0), (0, 0, 1)\}$, both vectors in the set are minimal vectors.

Given data on product characteristics x and discrete choices y for an infinite number of consumers, we have proved that only the true distribution of utility functions could have generated the data. Alternatively, the researcher could have data on market shares across a large number of markets. We discuss allowing unobservables at both the market and consumer levels in Section 5.1.

Example. We now discuss how the result in Theorem 2 specializes to the common empirical specification of random coefficients entering linear indices in utilities from Section 2. The equivalent of Assumption 1 is that $v = (v_1, \ldots, v_J)$ and $r = (r_1, \ldots, r_J)$ have support on a product space (i.e., they all independently vary) and that r has large support. The equivalent of Assumption 2 is that the heterogeneous unobservable $\theta = (\beta, \epsilon_1, \ldots, \epsilon_J)$ is distributed independently of the observable product characteristics $x = (v, r)$. The equivalent of Assumption 3 is that utility to heterogeneous unobservable θ purchasing product j is $v_j'\beta + \epsilon_j + r_j$. The equivalent of Assumption 3 then implies the equivalent of Assumption 4, as $v_j'\beta + \epsilon_j + r_j$ is real analytic in v_j. Under these assumptions, Theorem 2 implies that the distribution G of $\theta = (\beta, \epsilon_1, \ldots, \epsilon_J)$ is identified in the class of distributions with unknown finite support on $\mathbb{R}^{\dim(\theta)}$. In other words, we identify

$$\left(N, \left(\left(\beta^n, \epsilon_1^n, \ldots, \epsilon_J^n\right), w_{\theta n}\right)_{n=1}^{N}\right),$$

where N is the identified number of support points, $\theta^n = (\beta^n, \epsilon_1^n, \ldots, \epsilon_J^n)$ is the location of the nth support point, and $w_{\theta n}$ is the weight on the nth support point.

4.1 Purchasing Multiple Products

The empirical paper Liu, Chintagunta and Zhu (2010) studies a choice situations where each discrete choice $j = 0, \ldots, J$ indexes a bundle of composite choices. For example, a consumer can purchase cable television separately ($j = 1$), purchase an internet connection separately ($j = 2$), purchase both cable television and an internet connection together as a bundle ($j = 3$), or purchase nothing, the outside good ($j = 0$). The goal in this situation is to distinguish between explanations for observed joint purchase: are consumers observed to buy cable television and an internet connection at the same time because those who watch lots of television also have a high preference for internet service, or is there some causal utility increase from consuming both television and internet service together? The goal is to distinguish unobserved heterogeneity in preferences for products, which may be correlated across products, from true complementarities.

In our notation, heterogeneity is just captured by a distribution $G(\theta)$ that gives positive correlation between the choice-specific utility functions $u^1(v)$, $u^2(v)$, and $u^3(v)$. True complementarities are measured by

$$\Delta(v) \equiv u^3(v) - \left(u^1(v) + u^2(v)\right).$$

If the utility for choice j is $u^j(v) - r_j$ and r_j is the price of j, then $\Delta(v)$ is the monetary value of complementarities to the consumer. $\Delta(v) > 0$ represents a positive benefit from joint consumption. As utility functions are random functions across the population, there is a distribution of complementarity functions $\Delta(v)$ implied by $G(\theta)$.

As we have already explored in Theorem 2, we can identify the joint distribution of heterogeneous unobservables θ, which means we can identify the distribution of complementarities as a function of the joint distribution $G(\theta)$, if prices r_j are bundle-specific. That theorem imposed no restrictions on the joint distribution of utility functions across (here) distinct bundles. Therefore, we can apply Theorem 2 if we observe different choice situations where the bundle is or is not aggressively priced relative to the singleton packages. This is the data scheme for Liu et al.: they observe different bundles of telecommunications services at different prices, across geographic markets.

5 Endogenous Product Characteristics

This section explores endogenous product characteristics in the multinomial choice model. Recalling the discussion of the multinomial choice model in Section 4, an endogeneity problem arises when a consumer's preferences as captured by the utility functions $u = \left(u^1, \ldots, u^J\right)$ are not distributed independently of some elements of the consumer's product characteristics (v, r). We will first consider a motivation arising purely from consumer-level unobservables and then discuss market-level unobservables below.

Endogeneity from consumer-level unobservables could arise if the product characteristics (v, r) that an consumer faces is partly "designed" on the basis of information related to the consumer's vector of utility functions $u = \left(u^1, \ldots, u^J\right)$. A classic example of this source of endogeneity arises in a principal-consumer relationship, in which the principal designs the menu of product characteristics (v, r) facing the consumer using information that is correlated with the consumer's preferences u but that is not observable by the econometrician. The principal, here equivalent to a multi-product monopolist, has incentives (rent extraction) to use all known information about the consumer to design the menu of product characteristics for the consumer. Therefore, the endogenous choice of a menu of product characteristics (v, r) by a principal with some knowledge of $u = \left(u^1, \ldots, u^J\right)$ will induce a statistical endogeneity problem.

In this section, we show how to address the endogeneity problem posed by endogenous product characteristics in multinomial choice by way of a system of equations. Essentially, the system jointly models the decisions of the principal and the consumer, and uses exogenous variation in the characteristics of the principal-consumer relationship to achieve identification. We extend the notation from Section 4 to fit the new system of equations into the notation for a generic economic model in Section 3. Given $\tilde{v} \in \mathbb{R}^M$, $v \in \mathbb{R}^K$, and $r \in \mathbb{R}^J$, a consumer has utility for choice j given by $u^j(\tilde{v}, v) + r_j$. We let the first M product characteristics facing the consumer be potentially endogenous, denoting these product characteristics by $\tilde{v} \in \mathbb{R}^M$. We label the remaining, known-to-be-exogenous product characteristics $v \in \mathbb{R}^K$. We also refer to the possibly endogenous product characteristics $\tilde{v}$ as the principal's control variables, as they are strategically set by the principal. A special case of this framework is where $M = J$ and there is one endogenous control variable per product. We still normalize the outside good's utility to 0.

We introduce a vector of instruments $z = (z_1, \ldots, z_M) \in Z \subseteq \mathbb{R}^M$ that are stochastically independent of the vector of utility functions u. In addition, the instruments shift the endogenous choice characteristics through the principal's optimal choice of product characteristics, what in terms of identification is the instrumental variables equation $\tilde{v} = h(v, z)$ for $z \in Z$, $v \in V$, and $h : V \times Z \to \mathbb{R}^M$.

Let us map this new model in the notation for the generic economic model in Section 3. A heterogenous unobservable θ now corresponds to a pair of vectors of functions $\theta = (u, h)$ consisting of a vector of J utility functions $u = \left(u^1, \ldots, u^J\right)$ and a vector of M instrumental variable equations $h = \left(h^1, \ldots, h^M\right)$. An observable and statistically exogenous economic environment x is now $x = (v, r, z)$, a vector comprised of K exogenous product characteristics in the vector v, J exogenous special regressors in the vector r, and M exogenous, excluded instruments in the vector z.

The model is such that for any statistically exogenous economic environment $x = (v, r, z)$, the observed dependent variables are the consumer's discrete choice j and the principal's choice of the endogenous product characteristic $\tilde{v}$. In notation,

$$(j, \tilde{v}) = f(x, \theta) = f((v, r, z), (u, h)) \equiv \left(\arg\max_{j \in \mathcal{Y}} \left\{ \left\{ u^j(h(v, z), v) + r_j \right\}_{j=1}^J, 0 \right\}, h(v, z) \right),$$

where the principal's choice of endogenous product characteristics $\tilde{v}$ and the consumer's choice of product j are linked through a recursive system. One can see that a heterogeneous unobservable $\theta = (u, h)$ indexes an entire principal-consumer relationship, where the principal's function h generating endogenous product characteristics is itself heterogeneous due in part to differing information sets, cost functions or preferences among principals.

Importantly, the joint distribution $G(\theta) = G(u, h)$ over heterogeneous unobservables allows the principal's vector of functions h to be stochastically dependent with the consumer's utility functions u. The principal can condition its policy h for choosing product characteristics on information related to the consumer's preferences u that is unobserved to the econometrician. The instruments z are most naturally interpreted as ingredients to the marginal costs of providing each endogenous product characteristic, although they could represent any observed characteristics of the principal that are excluded from the utilities of consumers, including observed dimensions of the principal's information set or any other demographic taste shifters for the principal.

Remark 5. Our approach to handling endogeneity in the multinomial choice model parallels the approach in Fox and Gandhi (2011) for the case of continuous outcomes. In that paper, we argue that allowing an infinite dimensional heterogeneous unobservable in the endogenous product characteristic function h in the system distinguishes this approach from the nonparametric control function literature, which requires that the unobservable in the endogenous product characteristic function be invertible (Altonji and Matzkin, 2005; Chesher, 2003; Imbens and Newey, 2009; Blundell and Matzkin, 2010). Allowing an infinite dimensional unobservable u in the demand equation distinguishes this approach from the literature on simultaneous equations, which typically assumes that there is only a scalar unobservable for each product, so that the unobservables can be uniquely recovered for each statistical observation (Brown, 1983; Roehrig, 1988; Benkard and Berry, 2006; Matzkin, 2008; Berry and Haile, 2012; Berry, Gandhi and Haile, Forthcoming). In the parametric discrete choice literature, Petrin and Train (2010) is an example of the control function approach and Berry, Levinsohn and Pakes (1995) is an example of the inversion approach. Fox and Gandhi (2011) show that this identification approach can be consistent with Bertrand-Nash equilibrium.

Remark 6. Each θ, each realization of the unobserved heterogeneity, has its own $h(v, z)$ function. Therefore, in part we seek to identify, through $G(\theta)$, a distribution of functions h for the endogenous product characteristics, with arguments the exogenous product characteristics v and excluded instruments z. A special case of our analysis would be when we restrict h to be a real analytic function known up to a finite vector of unobservables, and we allow those unobservables themselves to be heterogeneous with unknown finite support. In empirical applications, we believe researchers would assume h is a known function up to a finite vector of heterogeneous parameters.

Remark 7. We do not allow the special regressors r to be endogenous or enter the endogenous product characteristic function h. The r's could reflect information about consumer and product characteristics that is unobserved to principals and distributed independently of principals' cost functions. Alternatively, the r's can capture an observable consumer attribute, such as location, that the seller cannot use as a basis for price discrimination or that does not convey information on a consumer's utility function vector u. If we did allow the vector of J r_j's to be endogenous, we conjecture we would need J special regressors (large support but not a common sign) as instruments so that the vector r could be "moved around" by the vector of instruments over r's large support. Currently we require the vector of endogenous product characteristics $\tilde{v}$ to move only locally with instruments.

Remark 8. By assuming that $x = (v, r, z)$ is distributed independently of $\theta = (u, h)$, we are assuming that the process that matches principals to consumers is exogenous and only product characteristics are endogenous. Otherwise consumers with certain unobservable preferences may be more likely to match with principals with certain observable cost shifters, making z an invalid instrument. Extending our framework to deal with endogenous matching is an interesting area for future research. Nevertheless there are numerous applied settings that fit our current version of the model. Consider Einav, Jenkins and Levin (2012), where the principal is a subprime auto dealer and the the customers exogenously arrive and desire cars with certain characteristics $(v, \tilde{v}, r)$. The principal can choose contract terms $\tilde{v}$ such as the minimum down payment and the interest rate for each car j. Consumers have heterogeneous preferences over minimum down payments and interest rates, perhaps reflecting varying liquidity constraints.

The formal assumptions for this model follow.

Assumption 5. *Let $V \subset \mathbb{R}^K$, the support of v, be a non-empty rectangle. Let $Z \subset \mathbb{R}^M$ also have support equal to a non-empty rectangle. Let $x = (v, r, z) \in \mathcal{X} = V \times R_1 \times \cdots \times R_J \times Z$ where $R_j = \mathbb{R}$ for each $j \in J$.*

Assumption 6. *The heterogenous unobservable $\theta = (u, h)$ is distributed independently of the observed, exogenous information $x = (v, r, z)$.*

We list the following as an assumption to emphasize the quasi-linearity restriction.

Assumption 7. *The utility of a consumer with utility functions $u = \left(u^1, \ldots, u^J\right)$ purchasing product j is $u^j(\tilde{v}, v) + r_j$.*

Assumption 8. *The space of heterogenous unobservables Θ consists of pairs $\theta = (u, h)$ such that u and h are both vector-valued analytic functions where the derivative $D_z h(v, z)$ of each heterogeneous instrumental variables equation with respect to the instruments satisfies the following conditions:*

1. *the derivative $D_z h(z, v)$ exists in the interior of $V \times Z$, and is continuous in $(v, z) \in V \times Z$; and*
2. *the derivative $D_z h(z, v)$ with respect to z has full rank M for almost all (in the sense of Lebesgue measure) $z \in Z$.*

Such a full rank restriction is a formal way of saying that the instrument z is a locally powerful instrument almost everywhere. For any heterogeneous unobservable $\theta = (u, h) \in \Theta$, almost everywhere local variation in z can induce the endogenous product characteristics $(\tilde{v}_1, \ldots, \tilde{v}_J)$ to vary locally in a full rank way, holding the exogenous product characteristics v fixed. Thus fixing $v \in V$ and for almost all $z \in Z$, the local variation in $\tilde{v}$ induced by the local variation in z is not restricted to a lower dimensional subspace.

Theorem 3. *Under Assumptions 5, 6, 7, and 8, the distribution $G^0 \in \mathcal{G}$ of $(u, h) \in \Theta$ in the multinomial choice model with endogenous product characteristics is identified.*

Proof. We verify the property in Theorem 1. Thus we take an arbitrary finite set of heterogeneous unobservables $T \subset \Theta$ and seek to construct a singleton I-set. For any finite set of heterogeneous unobservables $T \subset \Theta$, we form a singleton I-set of the form

$$I^T_{(0,\tilde{v}),(v,r,z)} = \left\{(u,h) \in T \mid h(v,z) = \tilde{v} \text{ and } u^j\left(h\left(v,z\right),v\right) + r_j \leq 0\,\forall j \in \{1,\ldots J\}\right\},$$

where recall good 0 is the outside option that has a normalized utility of 0. The I-set corresponds to the set of heterogeneous unobservables whose instrumental variables equation yields $\tilde{v}$ at $x = (v,r,z)$ and who choose the outside good.

Let $T_1 = \{h \mid \exists\, u \text{ such that } (u,h) \in T\}$. That is, T_1 is the set of distinct instrumental variables equations that arise within the set of heterogeneous unobservables T. By the properties of real analytic functions in Appendix A, there exists a tie breaking point $(v,z) \in V \times Z$ (which without loss can be assumed to be an interior point) such that for any distinct functions h_i and h_j in T_1, $h_i(v,z) \neq h_j(v,z)$. Consider a point $\tilde{v}^*$ from the set of values $\{h(v,z) \mid h \in T_1\}$. By construction, $\tilde{v}^*$ is attained at a unique $h \in T_1$; a unique $h \in T_1$ satisfies $\tilde{v}^* = h(v,z)$. Let us denote this unique $h \in T_1$ as h_1. By finiteness of the number of heterogeneous unobservables in T_1 and the fact that each $h \in T_1$ is continuous, $h_1(t_1,t_2) \neq h(t_1,t_2)$ for all $h \in T_1$ with $h \neq h_1$ and all $(t_1,t_2) \in W \subseteq V \times Z$, where W is a sufficiently small open neighborhood containing (v,z). There are now two cases to consider.

In case 1, the set $T_2 = \{(u,h) \in T \mid h = h_1\}$ is a singleton, which contains the single θ that we denote as (u_1,h_1), where $u_1 = \left(u_1^1,\ldots,u_1^J\right)$ and $h_1 = \left(h_1^1,\ldots,h_1^M\right)$. Setting $r_j = -u_1^j\left(h_1\left(v,z\right),v\right)\,\forall\, j$ ensures that the total utility of the heterogeneous unobservable (u_1,h_1) for each choice is 0. Because the definition of $I^T_{(0,\tilde{v}),(v,r,z)}$ states that consumers break ties in favor of the outside good, then $I^T_{(0,\tilde{v}^*),(v,r,z)}$ is a singleton, namely a set consisting of only $(u_1,h_1) \in T$.

In case 2, we have that the set $T_3 = \{u \mid (u,h_1) \in T_2\}$ is not a singleton. Observe that by Assumption 8, we can find a $z^* \in Z$ such that $(v,z^*) \in W$ and the Jacobian $D_z h_1(v,z^*)$ has full rank M. Furthermore, by the existence, continuity and almost-everywhere-in-z full rank of the Jacobian matrix $D_z h_1(v,z)$ in Assumption 8, the Jacobian matrix $D_z h_1(t_1,t_2)$ has full rank M for all (t_1,t_2) in a sufficiently small ball $C \subseteq W$ containing (v,z^*). This is because full rank occurs when the matrix in question has a nonzero determinant and the determinant of a matrix is a continuous function of the elements of the matrix.

As a consequence of the Jacobian having full rank everywhere in C and the open mapping theorem, the change of variable mapping $(v,z) \mapsto (h(v,z),v)$ defined over C, which we denote by B, is an open mapping.[5] Thus the image $B(C)$ is an open set in $\mathbb{R}^{M+K}$. Now using the properties of real analytic functions in Appendix A and the fact that $B\left(C\right)$ is an open set, there exists $(v',z') \in C \subseteq V \times Z$ such that for all distinct vectors u_i and u_k in T_3, $u_i\left(h_1\left(v',z'\right),v'\right) \neq u_k\left(h_1\left(v',z'\right),v'\right)$. Find a minimal vector $\underline{u}\left(h_1\left(v',z'\right),v'\right)$, which by the argument in the proof of Theorem 2 exists although there may

[5]The matrix of partial derivatives of B is of the form $H = \begin{bmatrix} D_v h(v,z) & I_K \\ D_z h(v,z) & 0_{J,K} \end{bmatrix}$, where I_K is an identity matrix with K rows and $0_{J,K}$ is a matrix of all 0's with J rows and K columns. The matrix H is invertible because $D_z h(v,z)$ is invertible. Therefore, by the open-mapping theorem, $(v,z) \mapsto (h(v,z),v)$ is an open mapping.

be multiple minimal vectors. Set $r_j = -\underline{u}^j\left(h_1\left(v', z'\right), v'\right) \forall j$. Then the heterogeneous unobservable $(\underline{u}, h_1)$ has 0 total utility for all choices and all other heterogeneous unobservables in T_2 have positive utility for some choice. Then $I^T_{(0,\tilde{v}^*),(v,r,z)}$ is a singleton, namely a set consisting of only $(\underline{u}, h_1) \in T$. $\square$

5.1 Unobservables At Both the Market and Consumer Levels

A common situation in demand estimation is that the endogenous regressor is price and price is the same across all consumers in a market. In this case, a conventional assumption is that the unobservables that are correlated with price (market-level demand shocks, say) are independent of consumer-level unobservables reflecting individual heterogeneity. Statistical independence between unobservables is a special case of our framework, which does not impose such assumptions. Further, in a world with J competing products, the unobservables reflecting demand and supply for one product may be statistically dependent with the prices of all products.

To see this notationally, as before let $\theta = (u, h)$, where now by assumption the function h that generates the endogenous price is the same for all consumers in a market. The realized utility functions u are now statistically dependent with h because the market level price is correlated with market-level unobservables that enter the otherwise consumer level utility functions. This type of model can be estimated using either individual data on consumer purchases across markets or aggregate data on market shares across markets.

6 Nonparametric Estimation

This paper's identification arguments, based on the identification machinery for the generic economic model in Theorem 1, are not constructive. In other words, the proof of Theorem 1 does not show how to constructively trace out the distribution of the heterogeneous unobservables. However, now we argue that our results on identification do fundamentally lead to proving the consistency of a nonparametric estimator for a distribution of heterogeneous unobservables. Here we focus on the more empirically relevant case where Θ, the space of heterogeneous unobservables, is finite dimensional, indeed a subset of the reals. For estimation (not identification), we require that Θ be compact. For a multinomial choice problem, we require that each utility function $u^j\left(v; \beta_j\right)$ be known up to finite heterogeneous parameters β_j, where the complete heterogeneous unobservable is $\theta = \beta = (\beta_1, \ldots, \beta_J)$, the vector of all choice-specific heterogeneous parameters.

We use a computationally simple, nonparametric estimator for distributions of random coefficients due to Fox, Kim, Ryan and Bajari (2011) and Fox, Kim and Yang (2013). Return to the generic economic model model of Section 3, but now assume the researcher has data on M observations (y_i, x_i) for consumers $i = 1, \ldots, M$. We wish to estimate G using this finite sample. First, we discretize the space of dependent variables $\mathcal{Y}$ into J categories $A_1, \ldots, A_J$. Let $y_{i,j} = 1$ if observation

i's y_i is in the set A_j, and 0 otherwise. For computational simplicity, in an initial stage we pick a grid of S points $\theta^s \in \Theta$. In a finite sample, we then estimate the weights $\tilde{w}_s$ on each grid point θ^s via the linear probability model regression

$$y_{i,j} \approx \sum_{s=1}^{S} \tilde{w}_s 1\left[f\left(x_i, \theta^s\right) \in A_j\right] + e_{i,j} \, \forall\, j \in 1, \ldots J,$$

where the approximation sign $\approx$ indicates that the set of S grids points is chosen for estimation and may not be the true set of grid points and $e_{i,j}$ is the error term in the linear probability model, $e_{i,j} \equiv y_{i,j} - \Pr\left(A_j \mid x_i\right)$, which satisfies $E\left[e_{i,j} \mid x_i\right] = 0$ by the definition of a choice probability. Because the grid points θ^s are fixed in estimation, the unknown weights $\tilde{w}_s$ in the linear probability model can be estimated via linear regression, which is computationally quite simple. Typically, one would add the linear inequality constraints $\tilde{w}_s \geq 0 \, \forall\, s$ and $\sum_{s=1}^{S} \tilde{w}_s = 1$. These ensure that the estimated weights form a valid probability mass function. The entire computational procedure is then linear least squares subject to linear inequality and equality constraints. The optimization problem is globally convex and specialized routines are guaranteed to find the constrained global minimum of the objective function. The estimate of $G\left(\theta\right)$, the distribution of random coefficients, is then

$$\hat{G}_M\left(\theta\right) = \sum_{s=1}^{S} \tilde{w}_s 1\left[\theta^s \leq \theta\right].$$

The constraints ensure this estimator is a valid cumulative distribution function.

Example. Consider the simple multinomial choice example in Section 2. There, $x = (v_1, \ldots, v_J, r_1, \ldots, r_J)$ and the heterogeneous unobservable $\theta = (\beta, \epsilon_1, \ldots, \epsilon_J)$. In this case, A_j is simply the choice j and so $y_{i,j}$ equals 1 when $y_i = j$ and 0 otherwise. The researcher picks a grid of S points $\theta^s = (\beta^s, \epsilon_1^s, \ldots, \epsilon_J^s)$. This grid should be chosen to provide a good approximation to the true distribution of the additive errors and marginal utilities for the product characteristics. For each statistical observation i, there are J regression observations corresponding to the linear probability model regression equation for each of the J choices. The term $1\left[f\left(x_i, \theta^s\right) \in A_j\right]$ just asks whether the utility of choice j with heterogeneous unobservable θ^s, $v_{i,j}'\beta^s + \epsilon_j^s + r_{i,j}$, is higher than the utility of the other $J-1$ inside goods and the outside good. Computationally, the researcher regresses the outcome $y_{i,j}$ on a vector of 1's and 0's corresponding to whether each heterogeneous unobservable θ^s would purchase good j at product characteristics x_i. This gives an estimate of the joint distribution of $(\beta, \epsilon_1, \ldots, \epsilon_J)$.

This is a sieve estimator (Chen, 2007). As the sample size M grows, one typically uses finer grids of $S\left(M\right)$ points. Let $\Theta_{S(M)}$ be the entire grid of points chosen with M observations. Under conditions on the choice of $S\left(M\right)$ and other conditions, the estimator $\hat{G}_M\left(\theta\right)$ converges to the true $G^0\left(\theta\right)$. Let $\vec{y}_i = (y_{i,1}, \ldots, y_{i,J})$ and let $\vec{P}\left(x, G\right) = \left(P_1\left(x, G\right), \ldots, P_J\left(x, G\right)\right)$, where $P_j\left(x, G\right) =$

$\int 1\left[f(x,\theta)\in A_j\right]dG(\theta)$. Fox et al. (2013) prove the following consistency proposition, using our notation.

Proposition 1. *Let the following conditions hold.*

1. *Let $\mathcal{G}$ be a space of distribution functions on $\Theta \subset \mathbb{R}^{\dim(\theta)}$, where Θ is compact. $\mathcal{G}$ contains the true G^0.*
2. *Let $\{(y_i, x_i)\}_{i=1}^N$ be i.i.d.*
3. *Let θ be independently distributed from x.*
4. *Assume that G^0 is identified, meaning that for any $G^1 \neq G^0$, $G^1 \in \mathcal{G}$, we have $\vec{P}\left(x, G^0\right) \neq \vec{P}\left(x, G^1\right)$ for almost all $x \in \tilde{\mathcal{X}}$, where $\tilde{\mathcal{X}}$ is a subset of $\mathcal{X}$, the support of x, with positive probability.*
5. *The population least squares objective function $E\left[\|y - P(x,G)\|_E^2\right]$ is continuous on $\mathcal{G}$ in the weak topology, where $\|\cdot\|_E$ is the Euclidean norm.*
6. *Let Θ_S become dense in Θ as $S \to \infty$.*
7. *$\Theta_S \subseteq \Theta_{S+1} \subseteq \Theta$ for all $S \geq 1$.*
8. *$S(M) \to \infty$ as $M \to \infty$ and it satisfies $\frac{S(M)\log S(M)}{M} \to 0$ as $M \to \infty$.*

Then $d_{\mathrm{LP}}\left(\hat{G}_M, G^0\right) \xrightarrow{p} 0$, where d_{LP} is the Lévy-Prokhorov metric on the space of multivariate distributions.

The proposition states that the estimated cumulative distribution function converges to the true distribution function G^0 as the sample size gets large, in a particular metric on the space of distribution functions.[6] Thus, the computationally simple linear regression estimator is a consistent nonparametric estimator for the distribution of heterogeneous unobservables.

In Proposition 1, our Theorem 1 has maintained conditions 2 and 3. In this section, we previously assumed condition 1, compactness. Conditions 6–8 are conditions on the choice of the grid of points for estimation. Satisfying these conditions is up to the empirical researcher. Condition 5 can usually be satisfied if some of the regressors are continuous, as Fox et al. (2013) show in some detail.

In Proposition 1, condition 4 is about identification. It states that the population data are compatible with only the true distribution of random coefficients, G^0. This condition is what this paper

[6] The function d_{LP} denotes the Lévy-Prokhorov metric $d_{\mathrm{LP}}(\mu_1, \mu_2)$, where μ_1 and μ_2 are probability measures corresponding to the distributions G_1 and G_2. The Lévy-Prokhorov metric is defined as

$$d_{\mathrm{LP}}(\mu_1,\mu_2) = \inf\left\{\epsilon > 0 \mid \mu_1(C) \leq \mu_2(C^\epsilon) + \epsilon \,\text{and}\, \mu_2(C) \leq \mu_1(C^\epsilon) + \epsilon \,\text{for all Borel measurable}\, C \in \Theta\right\},$$

where C is some set of heterogeneous unobservables and $C^\epsilon = \{\theta \in \Theta \mid \exists \theta' \in C,\, d(\theta',\theta) < \epsilon\}$. The Lévy-Prokhorov metric is a metric, so that $d_{\mathrm{LP}}(\mu_1,\mu_2) = 0$ only when $\mu_1 = \mu_2$.

has been about proving for the multinomial choice model. Identification is the key, model-specific condition needed to prove that this estimator for the distribution of random coefficients is consistent. If the distribution is not identified, there is no reason to think this or any other nonparametric estimator for the distribution of random coefficients will be consistent. Thus, the identification arguments in this paper lead directly to a computationally simple, nonparametric estimator for the distribution of random coefficients.

Remark 9. The notion of identification in condition 4 of Proposition 1 is slightly stronger than our notion of identification in Section 3, as our earlier notion requires only one x, not a set $\tilde{\mathcal{X}}$ with positive measure, where $\vec{P}\left(x, G^0\right) \neq \vec{P}\left(x, G^1\right)$. However, in the multinomial choice model, there will typically be an open set of x's where the I-sets $I_{A,x}^T$ are the same for an open ball around any x. In other words, an open set of x's will have the same $\theta \in T$ taking actions in the set A. In this case, the stronger notion of identification in condition 4 of Proposition 1 follows from the weaker notion proved, using the sufficient property, in Theorem 1, and hence Theorems 2 and 3 for the multinomial choice model.

Remark 10. We emphasize that the grid of points Θ_S used in a finite sample for estimation is not the same as the finite if unknown grid of points in the support of G^0, the true distribution. Proposition 1 states that the estimator $\hat{G}_M$ is consistent no matter the true support of G^0. The true support could be continuous, discrete or mixed discrete and continuous. The main criterion the model needs to satisfy is identification, as in condition 4 of Proposition 1. In the current paper, we show identification for the multinomial choice model when the true support of G^0 is indeed finite. However, Proposition 1 states that this support is learned in estimation as $M \to \infty$, and is not picked by the researcher.

7 Conclusions

We study identification of the distributions of heterogeneous unobservables, including random coefficients, in nonlinear economic models. We focus on the case with cross sectional data. For a generic nonlinear model with heterogeneous unobservables, we provide some general machinery for identification. We then straightforwardly verify that this machinery applies to the multinomial choice model. Our machinery may be of some independent interest, as it can presumably be applied to a variety of nonlinear models. The key assumption behind our identification result is that the support of the heterogeneous unobservables is unknown and finite in some possibly infinite dimensional space.

We are motivated by applications to demand estimation, where we observe consumers facing different product characteristics. In particular, we are the first to prove that the distribution of random coefficients entering linear indices for utilities in a multinomial choice model is identified, under some model-specific assumptions, including having one regressor for each choice with large support and a common sign that enters utility additively separably. Our identification result extends to identifying the distribution of heterogeneity in models with multiple purchases, if each bundle has a separate price. We can distinguish true complementarities from correlation in the values for individual products. Our results on multinomial choice also extend to the case where some regressors are endogenous, either from consumer level or market-level unobservables.

Our identification results lead directly to showing the consistency of a computationally simple, nonparametric estimator for the distribution of heterogeneous unobservables. Therefore, our results are useful for researchers wishing to use flexible specifications in empirical work.

A Real Analytic Functions

This appendix shows that the space of all real analytic functions satisfies a key property assumed in Assumption 4.

Definition 2. Let $\mathcal{X}$ be a non-empty rectangle in $\mathbb{R}^k$. A function $g : \mathcal{X} \to \mathbb{R}$ is **real analytic** if, given any interior point $\xi \in \mathcal{X}$, there is a power series in $x - \xi$ that converges to $g(x)$ for all x in some neighborhood $U \subset \mathcal{X}$ of ξ.

Real analytic functions are infinitely differentiable.

Definition 3. If a function $g = (g_1, \ldots, g_m) : \mathcal{X} \to \mathbb{R}^m$ is such that each of its m component functions g_i is real analytic, then g is **a vector-valued real analytic** function.

A property of the space of real analytic functions is that for any two distinct real analytic functions $g, g' : \mathcal{X} \to \mathbb{R}$, and for any open, connected set $C \subseteq \mathcal{X}$, g and g' cannot agree on the whole of C: there must exist $x \in C$ for which $g(x) \neq g'(x)$ (Krantz and Parks, 2002, Corollary 1.2.6). This property can easily be seen to extend to the space of vector-valued real analytic functions.

Proposition 2. *For any finite set of vector-valued real analytic functions* $\{g_1, \ldots, g_n\}$ and for any open $C \subseteq \mathcal{X}$, *there exists a point* $x \in C$ *such that* $g_i(x) \neq g_j(x)$ *for any distinct* g_i *and* g_j *in* $\{g_1, \ldots, g_n\}$.

Proof. Consider any finite set of vector-valued real analytic functions $\{g_1, \ldots, g_n\}$. We show by induction on n that the property holds for any finite number of elements n. The base case $n = 2$ holds by the above property of scalar-valued real analytic functions. To see this, for any nonempty, open set $C \subseteq \mathcal{X}$, take any non-empty ball within C, which is connected. Further, the two vector-valued real analytic functions g_1 and g_2 differ so they must have a different scalar function in at least one of the scalar slots of the two vectors. Then apply the above property of scalar-valued real analytic functions to this slot of the two vectors and to the non-empty ball within C.

Continuing with the induction, assume that the proposition holds for $n-1$, and consider $\{g_1, \ldots, g_n\}$ and a nonempty, open set $C \subseteq \mathcal{X}$, which without loss we can take to be an open ball (C contains such a ball, and balls are connected). By the induction hypothesis, there exists a point $x \in C$ such that $g_i(x) \neq g_j(x)$ for any $g_i \neq g_j$ and $i, j \in \{1, \ldots, (n-1)\}$. By the facts that each g_i is continuous and the set of functions is finite, these inequalities are preserved in a small open ball $B_1 \subseteq C$ around x. Now consider the function g_n, and observe that by the result for $n = 2$, there exists an $x_1 \in B_1$ such that $g_n(x_1) \neq g_1(x_1)$. Furthermore, by continuity, this inequality is preserved in a small ball $B_2 \subseteq B_1$ containing x_1. Now repeat the argument, except comparing g_n with g_2, producing a ball $B_3 \subseteq B_2$, etc. At the end of the process, a non-empty ball $B_n \subseteq B$ is produced for which any $x \in B_n$ satisfies the proposition, i.e., $x \in B_n$ implies $g_i(x) \neq g_j(x)$ for any distinct g_i and g_j in $\{g_1, \ldots, g_n\}$. $\square$

References

Altonji, J.G. and R.L. Matzkin, "Cross section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica*, July 2005, *73* (4), 1053–1102.

Benkard, C.L. and S. Berry, "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on Brown (1983) and Roehrig (1988)," *Econometrica*, 2006, *74* (5), 1429–1440.

Beran, R., "Prediction in random coefficient regression," *Journal of Statistical Planning and Inference*, 1995, *43* (1-2), 205–213.

_ **and PW Millar**, "Minimum Distance Estimation in Random Coefficient Regression Models," *The Annals of Statistics*, 1994, *22* (4), 1976–1992.

Berry, Steven T., Amit Gandhi, and Philip A. Haile, "Connected Substitutes and the Invertibility of Demand," *Econometrica*, Forthcoming.

_ **and Philip A. Haile**, "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," 2010. Yale University working paper.

_ **and** _ , "Identification in Differentiated Products Markets Using Market Level Data," 2012. Yale University working paper.

_ **, James Levinsohn, and Ariel Pakes**, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Blundell, Richard and Rosa L. Matzkin, "Conditions for the existence of control functions in nonseparable simultaneous equations models," 2010. UCL working paper.

Boyd, JH and RE Mellman, "Effect of Fuel Economy Standards on the U. S. Automotive Market: An Hedonic Demand Analysis," *Transportation Research Part B*, 1980, *14* (5), 367–378.

Briesch, Richard A., Pradeep K. Chintagunta, and Rosa L. Matzkin, "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," *Journal of Business and Economic Statistics*, 2010, *28* (2), 291–307.

Brown, B.W., "The identification problem in systems nonlinear in the variables," *Econometrica*, 1983, *51* (1), 175–196.

Cardell, NS and FC Dunbar, "Measuring the societal impacts of automobile downsizing," *TRANSPORT. RES.*, 1980, *14* (5), 423–434.

Chen, Xiaohong, "Large Sample Sieve Estimation of Semi-Nonparametric Models," in "Handbook of Econometrics," Elsevier Science, 2007.

Chesher, Andrew, "Identification in nonseparable models," *Econometrica*, September 2003, *71* (5), 1405–1441.

Chiappori, Pierre-André and Ivana Komunjer, "On the Nonparametric Identification of Multiple Choice Models," 2009. Columbia University working paper.

Cramér, H. and H. Wold, "Some Theorems on Distribution Functions," *Journal of the London Mathematical Society*, 1936, *1* (4), 290.

Domencich, T.A. and D. McFadden, *Urban travel demand: a behavioral analysis: a Charles River Associates research study*, North-Holland, 1975.

Einav, Liran, Mark Jenkins, and Jonathan Levin, "Contract Pricing in Consumer Credit Markets," *Econometrica*, 2012, *80* (4), 1387–1432.

Fox, Jeremy T. and Amit Gandhi, "Identifying Demand with Multidimensional Unobservables: A Random Functions Approach," October 2011. University of Michigan working paper.

_ , Kyoo il Kim, and Chenyu Yang, "A Simple Nonparametric Approach to Estimating the Distribution of Random Coefficients in Structural Models," 2013. University of Michigan working paper.

_ , _ , Stephen Ryan, and Patrick Bajari, "A Simple Estimator for the Distribution of Random Coefficients," *Quantitative Economics*, 2011, *2*, 381–418.

_ , _ , _ , and _ , "The Random Coefficients Logit Model is Identified," *Journal of Econometrics*, 2012, *1662* (2), 204–212.

Gautier, E. and Y. Kitamura, "Nonparametric Estimation in Random Coefficients Binary Choice Models," *Econometrica*, 2013, *81* (2), 581–607.

Hausman, J. and D. Wise, "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 1978, *46* (2), 403–426.

Heckman, J.J. and R.J. Willis, "A beta-logistic model for the analysis of sequential labor force participation by married women," *The Journal of Political Economy*, 1977, *85* (1), 27–58.

Hoderlein, Stefan, "Endogenous Semiparametric Binary Choice Models with Heteroscedasticity," 2009. Boston College working paper.

_ , Jussi Klemelä, and Enno Mammen, "Analyzing the Random Coefficient Model Nonparametrically," *Econometric Theory*, 2010, *26* (3), 804–837.

Ichimura, H. and TS Thompson, "Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution," *Journal of Econometrics*, 1998, *86* (2), 269–295.

Imbens, Guido W. and Whitney K. Newey, "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, September 2009, *77* (5), 1481–1512.

Khan, Shakeeb and Elie Tamer, "Irregular Identification, Support Conditions and Inverse Weight Estimation," *Econometrica*, 2010, *78* (6), 2021–2042.

Krantz, Steve G. and Harold R. Parks, *A Primer on Real Analytic Functions*, second ed., Birkhäuser, 2002.

Lewbel, Arthur, "Semiparametric latent variable model estimation with endogenous or mismeasured regressors," *Econometrica*, 1998, *66* (1), 105–121.

_ , "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 2000, *97* (1), 145–177.

Liu, Hongju, Pradeep Chintagunta, and Ting Zhu, "Complementarities and the Demand for Home Broadband Internet Services," *Marketing Science*, 2010, *29* (4), 701–720. University of Connecticut working paper.

Liu, Jingou, "The minimum distance method in nonlinear random coefficient models," *Statistica Sinica*, 1996, *6*, 877–898.

Magnac, Thierry and Eric Maurin, "Identification and information in monotone binary models," *Journal of Econometrics*, 2007, *139* (1), 76–104.

Manski, Charles F., "Identification of Binary Response Models," *Journal of the American Statistical Association*, September 1988, *83* (403), 729–738.

_ , "Partial Identification of Counterfactual Choice Probabilities," *International Economic Review*, November 2007, *48* (4), 1393–1410.

Matzkin, Rosa L., "Heterogeneous Choice," in W. Newey and T. Persson, eds., *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society*, Cambridge University Press., 2007.

_ , "Identification in Nonparametric Simultaneous Equations Models," *Econometrica*, September 2008, *76* (5), 945–978.

McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*, 1973, pp. 105–142.

Petrin, Amil and Kenneth Train, "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, February 2010, *47* (1), 3–13.

Roehrig, C.S., "Conditions for identification in nonparametric and parametric models," *Econometrica*, 1988, pp. 433–447.