

MIT Open Access Articles

The random coefficients logit model is identified

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Fox, Jeremy T., Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari. "The Random Coefficients Logit Model Is Identified." *Journal of Econometrics* 166, no. 2 (February 2012): 204–212.

As Published: <http://dx.doi.org/10.1016/j.jeconom.2011.09.002>

Publisher: Elsevier

Persistent URL: <http://hdl.handle.net/1721.1/101264>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-NoDerivatives



The Random Coefficients Logit Model Is Identified

Patrick Bajari

University of Minnesota & NBER

Jeremy T. Fox

University of Michigan & NBER

Kyoo il Kim

University of Minnesota

Stephen P. Ryan

MIT & NBER*

First Version: October 2006

This Version: September 2010

Abstract

The random coefficients multinomial choice logit model, also known as the mixed logit, has been widely used in empirical choice analysis for the last thirty years. We prove that the distribution of random coefficients in the multinomial logit model is nonparametrically identified. Our approach requires variation in product characteristics only locally and does not rely on the special regressors with large supports used in related papers. One of our two identification arguments is constructive. Both approaches may be applied to other choice models with random coefficients.

*Fox thanks the National Science Foundation, the Olin Foundation, and the Stigler Center and Bajari thanks the National Science Foundation, various grants, for generous funding.

1 Introduction

One of the most commonly used models in applied choice analysis is the random coefficients logit model, also known as the mixed logit, which models choice between one of a finite number of competing alternatives. Domencich and McFadden (1975), Heckman and Willis (1977) and Hausman and Wise (1978) introduced flexible specifications for discrete choice models, while the random coefficients logit model was proposed by Boyd and Mellman (1980) and Cardell and Dunbar (1980). Currently, the random coefficients logit model is widely used to model consumer choice in environmental economics, industrial organization, marketing, public economics, transportation economics and other fields.

In the random coefficients logit, each consumer can choose between $j = 1, \dots, J$ mutually exclusive inside goods and one outside good (good 0). The exogenous variables for choice j are in the $K \times 1$ vector $x_j = (x_{j,1}, \dots, x_{j,K})'$. In the example of demand estimation, x_j might include the product characteristics, the price of good j and the interactions of product characteristics with the demographics of agent i . Potentially x_j can depend on i but we suppress it for transparency. In this paper, for some results we assume the support of x_j includes the 0 vector, which can occur by centering each element of x_j around constants common across the inside goods. For example, for each k we may redefine $\tilde{x}_{j,k} = x_{j,k} - E[x_{j,k}]$, where the mean is taken over the inside goods.¹ Let $x = (x'_1, \dots, x'_J)$ denote the stacked vector of all the x_j . Each consumer i has a preference parameter β_i , which is a vector of K marginal utilities that gives i 's preferences over the K product characteristics. There is also a homogeneous term for each choice j , denoted as α_j . Each α_j could be an intercept common to product j or a term that captures product characteristics without random coefficients (homogeneous coefficients), as in $\alpha_j = \alpha + w'_j \gamma_w$. Here α is a common intercept for the inside goods contributing to the utility of purchasing an inside good instead of the outside good, γ_w is a vector of parameters on the characteristics in the vector w_j , and $w = (w'_1, \dots, w'_J)$ denotes the stacked vector of other product characteristics.² Let α^J be the vector of the J α_j 's. Agent i 's utility for choice j is equal to

$$u_{i,j} = \alpha_j + x'_j \beta_i + \epsilon_{i,j}. \quad (1)$$

The outside good has a utility of $u_{i,0} = \epsilon_{i,0}$.³ The logit model is defined when the errors $\epsilon_{i,j}$ are i.i.d. across choices and each error has the type I extreme value distribution, which has a cumulative distribution function of $\exp(-\exp(-\epsilon_{i,j}))$. The random coefficients logit arises when β_i varies across the population, with unknown distribution $F(\beta_i)$. The unknown objects of estimation are the distribution $F(\beta_i)$ and the homogeneous coefficients (α, γ'_w) or α^J . Under the standard assumption that β_i is independent of x_j (we discuss endogeneity below), utility maximization leads

¹This location normalization for the covariates is not without loss of generality in a model, like the logit, where the additive error is treated parametrically. However, one normalization is just as arbitrary as any other.

²We do not require intercepts; $\alpha_j = 0 \forall j$ is one possibility. Later we argue our approach does not apply to the case where the intercept for the inside goods, α , is also a random coefficient.

³The type I extreme value distribution gives the scale normalization for utility values. The outside good's utility gives the location normalization.

to a choice probability for the good j ,

$$\Pr(j | x; F, \alpha^J) \equiv \int \frac{\exp(\alpha_j + x'_j \beta_i)}{1 + \sum_{j'=1}^J \exp(\alpha_{j'} + x'_{j'} \beta_i)} dF(\beta_i). \quad (2)$$

This specification is popular with empirical researchers because the resulting choice probabilities are relatively flexible. In terms of modeling own- and cross-price elasticities, the random coefficients logit model allows products with similar x 's to be closer substitutes, which the logit model without random coefficients does not allow.⁴

In this paper, we prove that the distribution $F(\beta)$ is nonparametrically identified, in the sense that the true $(F^0, \alpha^{J,0})$ is the only the pair (F, α^J) that solves $\Pr(j | x, w) \equiv \Pr(j | x; F, \alpha^J)$ in (2) for all j and (x, w) , where $\Pr(j | x, w)$ denotes the population choice probabilities. We first recover the homogeneous terms, α^J (or the fixed parameters α and γ_w) in (2). We then provide two identification arguments, one of which is constructive and the other of which is non-constructive. The non-constructive identification theorem leverages results from Hornik (1991) on function approximation and Stinchcombe and White (1998) on hypothesis testing. The theorem requires variation in x in only an open set.

Our other identification argument is constructive. We demonstrate how to iteratively find all moments of β , which is sufficient to identify the distribution F^0 within the class of distributions that are uniquely determined by all of their moments. This class is the set of probability distribution functions that satisfy Carleman's condition, which we review below.

Both the constructive and non-constructive proof strategies are not unique to the logit: they could be applied to identify the distribution of heterogeneity in many differentiable economic models where covariates enter as linear indices. We outline the main theorems using generic notation and verify their conditions for the multinomial logit model.

While one of our identification approaches is constructive, we do not recommend that empirical researchers adopt an analog estimator to our identification argument. Instead, we suggest empirical users adopt one of several nonparametric mixtures estimators available in the literature. Before this paper, no one had formally proved that these mixtures estimators consistently estimated the true F^0 in the random coefficients logit. This lack of a complete consistency proof arises because showing that the density F^0 is nonparametrically identified is a necessary component for any consistency proof for a nonparametric estimator of F^0 . We introduce a computationally simple, nonparametric sieve estimator for F^0 in Bajari, Fox, Kim and Ryan (2010) for general mixtures models. We prove that our estimator is consistent in the Lévy-Prokhorov metric on distributions under the maintained assumption that the model is identified.⁵ This identification theorem therefore completes our proof

⁴McFadden and Train (2000, Theorem 1) present an approximation theorem using the random coefficients logit as the approximating class, although the theorem requires great flexibility in the choice of the product characteristics x_j in the random coefficients estimation as a function of some smaller set of underlying true product characteristics. McFadden and Train do not study identification.

⁵Alternative nonparametric estimators include the Bayesian MCMC estimators in Rossi, Allenby, and McCulloch (2005) as well as Burda, Harding and Hausman (2008) and the EM algorithm used in Train (2008). These works

of consistency for the estimator of the random coefficients logit in Bajari et al.⁶

The proof of identification is comforting to empirical researchers. Prior to our theorem, it was not known whether variation in x was sufficient to identify the pair $(F^0, \alpha^{J,0})$. One possibility was that the normality assumptions typically imposed on F^0 were crucial to identification. We show that indeed the random coefficients logit model is identified, which provides a more solid econometric foundation for its application in applied microeconomics.

The paper is organized as follows: Section 2 discusses the related literature, Section 3 states the general non-constructive identification result, Section 4 states the general constructive identification result, Section 5 shows how the results apply to the random coefficients logit, Section 6 explores extensions, and Section 7 concludes.

2 Related Literature

There is a growing literature on the identification of binary and multinomial choice models with unobserved heterogeneity. These papers differ in the set of assumptions made in order to obtain identification results and also differ in the objects of interest in identification.

Ichimura and Thompson (1997) study the case of binary choice: one inside good ($J = 1$) and one outside good. This restriction makes their method inapplicable for most empirical applications of demand analysis, which study markets with two or more inside goods. Ichimura and Thompson identify the cumulative distribution function (CDF) of, in our notation, $(\beta, \epsilon_{i,1} - \epsilon_{i,0})$. They use a theorem due to Cramer and Wold (1934) and do not exploit the structure of the extreme value assumptions on the $\epsilon_{i,1}$ and $\epsilon_{i,0}$. Consequently, they need stronger assumptions: 1) a monotonicity assumption (sign restriction) on one of the K components of β ($\beta_{i,k} > 0 \forall i$) and 2) a full support assumption for all K elements of $x_{i,1}$. Similar assumptions will appear in many of the papers below. We will refer to a regressor whose (random) coefficient has a sign restriction and that has full support as a “special regressor.” Gautier and Kitamura (2009) provide a computationally simple estimator and some alternative identification arguments for the same binary choice model as Ichimura and Thompson. To our knowledge, no one has generalized Ichimura and Thompson to the case of multinomial choice.

Lewbel (2000) provides an identification argument that relies on a large-support special regressor, but which allows for discrete elements of x_j , relaxes the independence between x and ϵ , and does not rely on distributional assumptions of the error term.⁷ Lewbel’s approach identifies the means of the

do not discuss consistency or identification. However, the identification theorem here would be a building block to proving the consistency of the estimates of (F, α^J) in these other procedures.

⁶Strictly speaking, the consistency Theorem 3.2 in Bajari et al (2010) applies to models without homogeneous parameters, such as the intercepts α^J . However, in Section 5 of that paper we discuss an extension to nonlinear least squares where location and scale parameters (homogeneous parameters) are also estimated. A previous draft of Bajari et al included the parametric verification of the regularity conditions for nonlinear least squares for models with both homogeneous parameters α and a distribution of heterogeneous parameters β .

⁷As a referee points out, the sign restriction or the monotonicity assumption for a special regressor may not be too restrictive because it is testable and the sign can be determined from a pre-model analysis as described in Lewbel (2000) and his other related work. Magnac and Maurin (2007) and Kahn and Tamer (2010), however, argue that

random coefficients and the conditional distribution of the composite error $x'_j(\beta_i - E[\beta_i | x]) + \epsilon_{i,j}$. However, identifying the distribution of the composite error does not identify the joint distribution of the random coefficients, which is the key object of interest in our paper.

Briesch, Chintagunta and Matzkin (2009) study the identification of a discrete choice model where the payoff to choice j is $V(j, z_j, s_i, \omega_i) - r_j + \epsilon_{i,j}$, where V is an unknown, nonparametric function common to all consumers, z_j are observed product characteristics, s_i are observed consumer characteristics, r_j is a large support special regressor with a sign restriction as in Ichimura and Thompson, $\epsilon_{i,j}$ is an additive error and ω_i is a scalar unobservable that enters the utility functions for all J choices. There are a variety of other restrictions. Their model does not nest the random coefficients logit. Matzkin (2007, page 101) extends these results to the utility function for choice j of $r_j + \sum_{k=1}^K m_k(z_{j,k}, \omega_{i,j,k}) + \epsilon_{i,j}$, where each function $m_k(\cdot, \cdot)$ is treated nonparametrically, each choice j has its own random coefficients $\omega_{i,j} = (\omega_{i,j,1}, \dots, \omega_{i,j,K})$, the random coefficients are independent across choices, and r_j is a special regressor. The special regressor and the independence of random coefficients across choices mean that the standard random coefficients logit model is not nested in the formulation of Matzkin.

Subsequent to the circulation of our constructive identification theorem, Berry and Haile (2010) and Fox and Gandhi (2009) introduced identification arguments for multinomial choice models without the type I extreme value distribution or additive errors. Both Berry and Haile and Fox and Gandhi need a monotonicity assumption on one of the K components of β ($\beta_{i,k} > 0 \forall i$) and (for point instead of set identification) a full support assumption on the corresponding k -th component $x_{j,k}$, for all choices $j \in J$. Berry and Haile identify the conditional-on- x_i distribution of utility values $G(u_{i,0}, u_{i,1}, \dots, u_{i,J} | x)$ and not $F(\beta)$. Knowledge of the full structural model, in the logit case $F(\beta)$, is necessary for welfare analysis, for example to construct the distribution of welfare gains between choice situations x^1 and x^2 , or some aggregation of welfare gains over individuals $H(\Delta_i u | x^1, x^2)$, where

$$\Delta_i u = \max_{j \in J \cup \{0\}} u_{i,j}(x^1) - \max_{j \in J \cup \{0\}} u_{i,j}(x^2),$$

where $u_{i,j}(x^l)$ is just the realized utility value (1) for $x^l = (x^l_1, \dots, x^l_J)$. Fox and Gandhi do identify the full structural model, in that they identify a distribution D over J utility functions (not utility values) of x , as in $D(u_{i,1}(x), \dots, u_{i,J}(x))$, where $u_{i,1}(x)$ is a complete function that describes utility values for choice j at all x . Again, like all results other than ours, Fox and Gandhi rely on monotonicity and large support assumptions for a special regressor.

Compared to this other literature, our main distinguishing feature is that we exploit the logit distributional assumptions on the $\epsilon_{i,j}$. This corresponds to empirical practice: the random coefficients logit is a popular specification in applied work. Our results contribute to the literature by demonstrating that large support and monotonicity restrictions are not required for identification if the logit error structure is used, as is common in empirical work. In our opinion, the main con-

identification using the large support condition and a conditional mean restriction critically relies on the tail behavior of the distribution of the special regressor, as is also the case in “identification at infinity” in selection models (Andrews and Schafgans 1998).

cern with the special regressor assumption is the requirement for large support. Large supports are sometimes but not often found in typical datasets used in demand estimation. Price may be a special regressor; certainly the assumption of monotonicity on price is rarely controversial. In an output market, prices are some markup over cost, and cost rarely moves more than a factor of, say, five (think oil price fluctuations). As we show, the parametric assumption on the distribution of the choice-specific errors does away with the need for large support assumptions. The entire distribution of random coefficients can be identified using only local variation in characteristics.⁸

Our two identification approaches may be applied to choice models with random coefficients other than the logit, as we describe below. Our paper, like many of the papers in the literature, focuses on continuous covariates in x .⁹ All arguments can be made conditional on the values of discrete covariates, but we do not explore identifying a distribution of random coefficients on discrete covariates. No other paper has identified such a distribution, either. Therefore, our results do not allow the random coefficients logit to be a flexible error components model for parameterizing the correlation between products grouped into various nests. Unlike the subset of the literature that is nonparametric on the contribution of x to the utility of each choice, we follow Ichimura and Thompson (1997) and the widespread empirical literature using the random coefficients logit and focus on the linear index $x'_j\beta_i$.

3 Non-Constructive Identification

To deliver the key idea behind our identification strategy, we shall first consider an abstract model that includes the random coefficients logit with fixed intercepts α as a special case. We present a non-constructive identification theorem and, in a later section, turn to constructive identification.

The econometrician observes $K < \infty$ covariates $x = (x_1, \dots, x_K)'$ and the probability of some discrete outcome, $P(x)$. Here $P(x)$ denotes the conditional choice probability of a particular outcome. For a model with a more complex outcome (including a continuous outcome y), we can always consider whether some event, say $y < \frac{1}{2}$, happened or did not happen. For a multinomial choice, the event could be picking choice j or picking any other choice. $P(x)$ is the probability of the event happening. The independent variable x is independent of β .

Let $g(\alpha, x'\beta)$ be the probability of an agent with characteristics β taking the action. In our framework, the researcher specifies $g(\alpha, x'\beta)$. A special case is $g(\alpha, x'\beta) = g(\alpha + x'\beta)$. Our goal is

⁸In subsequent and at present in-progress work, Chiappori and Komunjer (2009) present preliminary results for achieving the same weakening of conditions on special regressors without parametric assumptions in the multinomial choice model.

⁹For the case of binary choice, Lewbel (2000) allows for discrete covariates other than in the special regressor and shows the identification of, in our notation, the distribution of $x'_j(\beta - E[\beta|x]) + \epsilon_{i,j}$ conditional on x . He does not explore identification of an unconditional joint distribution $F(\beta)$. No paper has identified the distribution of β if some elements of x are discrete. Knowledge of the unconditional distribution $F(\beta)$ is necessary for some uses of structural models, including prediction of demand and x 's not in the support of the original data (the new goods problem).

to identify the distribution function $F(\beta)$ in the equation

$$P(x) = P(x, F, \alpha) \equiv \int g(\alpha, x'\beta) dF(\beta). \quad (3)$$

Identification means that a unique (F, α) solves this equation for all x . This is the definition of identification used in the statistics literature (Teicher 1963).

We prove the main theoretical results for the case where the true values α^0 of the homogeneous parameters α are identified in a first stage using an auxiliary argument, so we can write $g(x'\beta) = g(\alpha^0, x'\beta)$ with an abuse of notation. Typically, this auxiliary argument will involve the point $x = 0$, as then (3) becomes $P(0) = P(0, F, \alpha) = g(\alpha, 0)$ and in some models we can set $\alpha^0 = g^{-1}(P(0))$. The point $x = 0$ will otherwise not be needed to be in the support of x for the general result. We rewrite the model as

$$P(x, F) \equiv \int g(x'\beta) dF(\beta). \quad (4)$$

Let $\mathcal{B} \subseteq \mathbb{R}^K$ be the support of the random coefficients and let $\mathcal{F}(\mathcal{B})$ be the set of all distributions on that support. Let $\mathcal{X} \subseteq \mathbb{R}^K$ be the support of the covariates. Let F^0 be the true distribution. Then we have $P(x) = P(x, F^0)$ in the support of x , \mathcal{X} .

Definition 1. The distribution $F^0 \in \mathcal{F}(\mathcal{B})$ is **uniformly identified over choices of $(\mathcal{B}, \mathcal{X}_0)$** if for any $F^1 \in \mathcal{F}(\mathcal{B})$, $F^1 \neq F^0$, there exists $\mathcal{X}_0^1 \subset \mathcal{X}_0$ such that $P(x, F^0) - P(x, F^1) \neq 0$ for all $x \in \mathcal{X}_0^1$ for any choices of the support of random coefficients \mathcal{B} and the subset of the support of covariates $\mathcal{X}_0 \subset \mathcal{X}$, where \mathcal{B} is compact and \mathcal{X}_0 is a nonempty open set.¹⁰

We do not assume the researcher knows \mathcal{B} , other than that it is compact. The “nonempty interior” assumption below does rule out covariates with discrete support, as we discussed earlier.¹¹ For explicitness, we emphasize the assumption on the covariates implied in the previous definition.

Assumption 2.

1. *The support of the independent variables, $\mathcal{X} \subseteq \mathbb{R}^K$ includes a nonempty open set.*
2. *Let all elements of x be continuous.*

Assumption 2.1 will be violated if x includes higher order terms of x_k 's or interactions of x_k 's (e.g., $x_2 = x_1^2$ or $x_3 = x_2 \cdot x_1$) in its elements. Therefore our identification results do not allow for those terms in the model.

Definition 3. The function $g(z)$ is **real analytic at c** whenever it can be represented as a convergent power series, $g(z) = \sum_{d=0}^{\infty} a_d (z - c)^d$, for a domain of convergence around c . The function $g(z)$ is **real analytic** on an open set \mathcal{Z} if it is real analytic at all arguments $z \in \mathcal{Z}$.

¹⁰In this definition the set \mathcal{X}_0^1 can vary depending on alternative distribution F^1 . When the measure of the set \mathcal{X}_0^1 is strictly positive, identification holds with a positive probability.

¹¹All arguments in this paper can be made conditional on the values of discrete characteristics. Otherwise and as a referee suggests, one may take a set identification approach when x takes values on a fine grid. However, this is beyond the scope of this paper.

Our non-constructive identification theorem follows.

Theorem 4. *Assume that the vector α^0 is identified using some auxiliary argument and let Assumption 2 hold. Let \mathcal{B} be compact and \mathcal{X}_0 be a nonempty open subset of \mathcal{X} . The distribution $F^0 \in \mathcal{F}(\mathcal{B})$ is **uniformly identified over choices of $(\mathcal{B}, \mathcal{X}_0)$** if the function $g(\cdot)$ is real analytic, nonconstant and satisfies $g(0) \neq 0$.*

3.1 Lemmas That Factor Into the Proof of Theorem 4

The proof of Theorem 4 is a consequence of two lemmas. We state the lemmas here for those interested in the logic behind the identification theorem.

We say that $F^0 \in \mathcal{F}(\mathcal{B})$ is uniformly identified over choices of \mathcal{B} , holding a set of values of x , \mathcal{T}_0 , fixed, whenever $P(x, F^0) - P(x, F^1) \neq 0$ for all $x \in \mathcal{T}_0^1 \subset \mathcal{T}_0$ for any $F^1 \in \mathcal{F}(\mathcal{B})$, $F^1 \neq F^0$ and for any compact choice for the support of the random coefficients, \mathcal{B} . In this definition \mathcal{T}_0 does not have to be included in the support of x . $P(x, F^0)$ and $P(x, F^1)$ for $x \in \mathcal{T}_0 \not\subseteq \mathcal{X}$ are well defined from the model in (4) and then we have $P(x) = P(x, F^0)$ only for $x \in \mathcal{X}$ if F^0 is the true distribution.

Lemma 5. *Assume that the vector α^0 is identified using some auxiliary argument. Let $g(\cdot)$ be real analytic and let a set of x , \mathcal{T} , contain a nonempty open set. The distribution $F^0 \in \mathcal{F}(\mathcal{B})$ is uniformly identified over choices of $(\mathcal{B}, \mathcal{T}_0)$ with nonempty open sets $\mathcal{T}_0 \subset \mathcal{T}$ if and only if $F^0 \in \mathcal{F}(\mathcal{B})$ is uniformly identified over choices of \mathcal{B} , for at least one fixed $\mathcal{T}_0 \subseteq \mathcal{T}$.*

The lemma and its short proof (in our appendix for completeness) are inspired by Theorem 3.8 in Stinchcombe and White (1998), a paper on hypothesis testing. The content of Lemma 5 is that identification for any choice of nonempty open set \mathcal{T}_0 automatically holds if identification is checked for one, likely convenient choice of \mathcal{T}_0 . The most convenient choice of \mathcal{T}_0 is the one with the widest variation, or $\mathcal{T}_0 = \mathcal{T} = \mathbb{R}^K$.

Lemma 6. *Assume that the vector α^0 is identified using some auxiliary argument. Let $g(\cdot)$ be bounded and nonconstant and satisfy $g(0) \neq 0$. Then the distribution $F^0 \in \mathcal{F}(\mathcal{B})$ is uniformly identified over choices of \mathcal{B} for the choice $\mathcal{T} = \mathbb{R}^K$.*

Lemmas 5 and 6 together imply Theorem 4. Lemma 6 shows the model (4) identifies F^0 against any $F^1 \neq F^0$ with $\mathcal{T}_0 = \mathcal{T} = \mathbb{R}^K$ and Lemma 5 says then identification should hold with any nonempty open set \mathcal{T}_0 . Therefore, we conclude that the F^0 that solves $P(x) = P(x, F)$ in (4) is identified with any nonempty open set $\mathcal{X}_0 \subset \mathcal{X}$, noting that we have $P(x) = P(x, F^0)$ in the support of x . Lemma 5 shows the role of $g(\cdot)$ being real analytic for identification: it removes the full support condition.

Lemma 6 follows from the proof of Theorem 5 in Hornik (1991), a paper on functional approximation using a function that takes a linear index. Other than the requirement that $g(0) \neq 0$, Lemma 6 would be exactly the same as the statement of Theorem 5 of Hornik (using different notation) if our model was $g(x_1'\beta + x_2)$, where in a study of identification x_2 is a special regressor with large support and known sign. However, an inspection of Hornik's proof shows x_2 is only used

to establish that $g(x_2) \neq 0$ for some $x_2 \in \mathbb{R}$ (because g is nonconstant), which we replace with the requirement that $g(0) \neq 0$. We do not need data on $x = 0$ to establish that $g(0) \neq 0$; indeed $P(0, F) = g(0)$ is known from the choice of g (and the identification of α), as it is a trivial function of F . As the proof of Theorem 5 in Hornik is long, we do not include it here. To our knowledge, no other paper has linked the result on function approximation in Hornik to the identification of distributions of heterogeneity.

4 An Alternative, Constructive Approach

The previous identification approach is fairly general but is not constructive. In other words, the identification argument implies that $P(x, F^0) - P(x, F^1) \neq 0$ for a set of x with positive probability¹² but does not give a procedure to construct F^0 . Here we give such a constructive identification argument. The downside is that an open set around the point $x = 0$ will play a special role in the identification of the distribution of random coefficients, while the point $x = 0$ only possibly played the role of identifying homogeneous parameters α in the prior, non-constructive argument. On the other hand, compactness of the parameter space \mathcal{B} will not be imposed, as it was in the non-constructive argument. Indeed, \mathcal{B} can equal \mathbb{R}^K . The function $g(\cdot)$ will not need to be real analytic. As before, the homogeneous parameters α can be identified in a first stage.

Assumption 7. *The absolute moments of $F(\beta)$, given by $m_l = \int \|\beta\|^l dF(\beta)$, are finite for $l \geq 1$ and satisfy the Carleman condition: $\sum_{l \geq 1} m_l^{-1/l} = \infty$.*

A distribution F satisfying the Carleman condition is uniquely determined by its moments (Shohat and Tamarkin 1943, p. 19). The Carleman condition is weaker than requiring the moment generating function to exist. The Carleman condition gives uniqueness for distributions with unrestricted support. If the support of F is known and compact, uniqueness follows without the Carleman condition. Let $g^{(l)}(c)$ be the l -th derivative of $g(c)$ evaluated at c .

Assumption 8.

1. $g(c)$ is infinitely differentiable on an open set $\mathcal{C} \subset \mathbb{R}$ that includes $c = 0$.
2. $g^{(l)}(0)$ is nonzero and finite for all $l \geq 1$.

For a (non-probability) example, if $g(c) = D \cdot \exp(c)$, then Assumption 8 is satisfied because $g^{(l)}(0) = D$ for all l . If $g(\cdot)$ is a polynomial function of any finite degree, g does not satisfy the condition because its derivative becomes zero at a certain point. For polynomials, we identify the distribution of β up to the v -th moment, where v is the order of the polynomial function.

The key limitation of the constructive identification theorem is its reliance on regressor variation in an open set around the point $x = 0$.

¹²To be precise this is because identification holds on a set with nonempty interior and we assume this set has positive measure.

Assumption 9. *The support \mathcal{X} contains an open set surrounding $x = 0$.*

Recall that, in a multinomial choice model, the point $x = 0$ could arise via re-centering.

The constructive identification argument is quite simple. The $P(x) = P(x, F^0)$ are the observed choice probabilities in the data. We illustrate the argument for the special case where $K = 2$ and so $x'\beta = x_1\beta_1 + x_2\beta_2$. At $x_1 = x_2 = 0$,

$$\left. \frac{\partial P(x, F^0)}{\partial x_1} \right|_{x=0} = g^{(1)}(\alpha) \int \beta_1 dF^0(\beta) = g^{(1)}(\alpha) E[\beta_1],$$

where β_1 arises from the chain rule and the expression identifies the mean of β_1 , because $P(x, F^0)$ is in the data and $g^{(1)}(\alpha)$ is a known constant that does not depend on β . Likewise, $\left. \frac{\partial P(x, F^0)}{\partial x_2} \right|_{x=0} / g^{(1)}(\alpha)$ equals $E[\beta_2]$, $\left. \frac{\partial^2 P(x, F^0)}{\partial x_1 \partial x_2} \right|_{x=0} / g^{(2)}(\alpha)$ equals $E[\beta_1\beta_2]$, and $\left. \frac{\partial^2 P(x, F^0)}{\partial x_1^2} \right|_{x=0} / g^{(2)}(\alpha)$ equals $E[\beta_1^2]$. Additional derivatives will identify the other moments of $\beta = (\beta_1, \beta_2)$. We make no assumption that the components β_1 and β_2 are independently distributed; F^0 is an unrestricted joint distribution.

Theorem 10. *Assume that the vector α^0 is identified using some auxiliary argument. Suppose Assumptions 2 and 9 hold.*

- *Suppose Assumptions 7 and 8 also hold. Then the true F^0 is identified.*
- *Assume the first L derivatives of $g(c)$ with respect to c are nonzero when evaluated at the scalar argument $c = 0$. Then all moments of β up to order L (including cross moments) are identified.*

The proof is in the appendix. Note the approach's simplicity: we need only to check for non-zero derivatives of $g(c)$ at $c = 0$. This technique can be applied to show identification of many differentiable economic models that use linear indices $x'\beta$. The approach is constructive: if $g^{(2)}(0) \neq 0$, we can identify all own second moments and all cross-partial moments between two random coefficients. If only the first 100 derivatives of $g(c)$ at $c = 0$ are nonzero, then we identify at least the first 100 moments of the random coefficients.

5 Identification of the Random Coefficients Logit Model

5.1 Homogeneous Parameters for the Logit

Our leading example of a mixtures model is the random coefficients logit as outlined in the introduction. We first show we can identify the homogeneous terms, α_j^0 for all j . Consider the point $x = 0$. Algebra shows that

$$\log \Pr(j | x = 0; F^0, \alpha^{J,0}) - \log \Pr(0 | x = 0; F^0, \alpha^{J,0}) = \alpha_j^0 \forall j = 1, \dots, J,$$

where $\Pr(j | x = 0; F^0, \alpha^{j,0})$ is identified from the data for all $j = 0, 1, \dots, J$. The vector $\alpha^{j,0}$ is identified at the point $x = 0$.

Following this, we can show the parameters (α, γ'_w) in $\alpha_j = \alpha + w'_j \gamma_w$ are identified. Let $\Pr(j | x, w) = \Pr(j | x, w; F^0, \alpha^0, \gamma_w^0)$ be the choice probability for product j when the regressors are x and w . The population linear regression of $\log \Pr(j | x = 0, w; F^0, \alpha^0, \gamma_w^0) - \log \Pr(0 | x = 0, w; F^0, \alpha^0, \gamma_w^0)$ on a constant and the vector w_j will identify the constant term α^0 and the homogeneous coefficients γ_w^0 as long as the density of $w = (w'_1, \dots, w'_J)$ is strictly positive on its support at $x = 0$ ¹³ and $E \left[(1, w'_j)'(1, w'_j) | x = 0 \right]$ is nonsingular. Thus, the homogeneous terms are identified from differences in market shares when all products are evaluated at $x = 0$. The identification of F^0 below can proceed at any given value of $w = \tilde{w}$ in the support of w as long as $x = 0$ is allowed at \tilde{w} .¹⁴ We collect these conditions for the identification of (α^0, γ_w^0) .

Assumption 11. *The density of $w = (w'_1, \dots, w'_J)$ is strictly positive on its support at $x = 0$ and $E \left[(1, w'_j)'(1, w'_j) | x = 0 \right]$ is nonsingular.*

5.2 Non-Constructive Identification for the Logit

Using some duplication of notation, we can fit the mixed logit model into the mixtures framework by defining the logit choice probabilities for some particular choice j as¹⁵

$$g_j(\alpha^J, x'_1\beta, \dots, x'_J\beta) = \frac{\exp(\alpha_j + x'_j\beta)}{1 + \sum_{j'=1}^J \exp(\alpha_{j'} + x'_{j'}\beta)}.$$

Let $x_{j'} = 0$ for all $j' \neq j$. With one outside good and J inside goods, the choice probability of alternative j given β is

$$g_j(\alpha^J, 0, \dots, x'_j\beta, \dots, 0) = \frac{\exp(x'_j\beta)}{\exp(-\alpha_j) + \sum_{j' \neq j} \exp(\alpha_{j'} - \alpha_j) + \exp(x'_j\beta)}.$$

Then we obtain the integrated choice probability

$$P_j(0, \dots, x_j, \dots, 0, F, \alpha^J) = \int g_j(\alpha^J, 0, \dots, x'_j\beta, \dots, 0) dF(\beta), \quad (5)$$

where $P_j(0, \dots, x_j, \dots, 0, F, \alpha^J)$ denotes the conditional choice probability of the good j at $x = (0', \dots, x'_j, \dots, 0')$ and α^J .

¹³If some elements of w_j are discrete, this is a density with respect to the counting measure for the discrete elements.

¹⁴Assume that x and w jointly have product support. The identification argument for (α^0, γ_w^0) directly relies on a set of (w, x) of measure zero only because of the focus on the point $x_1 = \dots = x_J = 0$, not because of any additional restrictions on w .

¹⁵When there is no outside good, the same approach is taken using a normalization with respect to a particular good, say good 1. Let $\tilde{\alpha}_j = \alpha_j - \alpha_1$ and $\tilde{x}_j = x_j - x_1$. Then the model becomes $g_j(\tilde{\alpha}^{J-1}, \tilde{x}'_2\beta, \dots, \tilde{x}'_J\beta) = \frac{\exp(\tilde{\alpha}_j + \tilde{x}'_j\beta)}{1 + \sum_{j'=2}^J \exp(\tilde{\alpha}_{j'} + \tilde{x}'_{j'}\beta)}$.

Define $A_j(\alpha^J) = \exp(-\alpha_j) + \sum_{j' \neq j} \exp(\alpha_{j'} - \alpha_j)$ and, in another duplication of notation, $g_j(\alpha^J, c) = g_j(A_j(\alpha^J), c) = \frac{\exp(c)}{A_j(\alpha^J) + \exp(c)}$, which is a function of a single argument c given A_j . Therefore, the formulation of (5) is a special case of (3), where we take $P(x_j, F, \alpha) = P_j(0, \dots, x_j, \dots, 0, F, \alpha^J)$ and $g(\alpha^J, x'_j \beta) = g_j(\alpha^J, x'_j \beta)$. The choice j identification focuses on is arbitrary.

First we state non-constructive identification theorem.

Assumption 12. *The support of x , \mathcal{X} contains $x = 0$, but not necessarily an open set surrounding it. Further, the support contains a nonempty open set of points (open in \mathbb{R}^K) of the form $(x'_1, \dots, x'_{j-1}, x'_j, x'_{j+1}, \dots, x'_J) = (0', \dots, 0', x'_j, 0', \dots, 0')$.*

The identification approach needs $x = 0$ to be in the support of x so that the homogeneous parameters are identified and so that identification can exclusively focus on variation in the characteristics of choice j . The point $x = 0$ otherwise plays no role in the identification of the distribution of random coefficients. In other words, identification of the random coefficients can come from a small open set of x_j values far from $x_j = 0$. The difference between Assumptions 12 and 9 is subtle, but the role of the point $x = 0$ in identification of F^0 is quite different in the constructive and non-constructive identification approaches.

Theorem 13. *Let the true model be the multinomial logit and let Assumptions 2, 11 and 12 hold. The homogeneous parameters $\alpha^{J,0}$ are identified. Also, the distribution $F^0 \in \mathcal{F}(\mathcal{B})$ is uniformly identified over choices of $(\mathcal{B}, \mathcal{X}_0^K)$, where $\mathcal{X}_0^K \subseteq \mathbb{R}^K$ are nonempty open subsets of the space of characteristics of one particular product, \mathcal{X}^K .*

Identification holds in any nonempty open set of product characteristics satisfying the conditions in the theorem. This theorem is a specialization of Theorem 4 to the multinomial logit. The logit $g_j(\alpha^J, x'_j \beta)$ is nonzero at all $x_j \in \mathcal{X}^K$ for finite α^J , is bounded, and is real analytic. Real analyticity holds because the function $\exp(\cdot)$ is real analytic and the function $g_j(\alpha^J, c)$ is formed by the addition and division of never zero real analytic functions, and so is itself real analytic (Krantz and Parks 2002). Thus, the distribution of random coefficients in the multinomial logit is non-constructively identified.

5.3 Constructive Identification

Let $g^{(l)}(A(\alpha^J), 0)$ be the l -th derivative of $g_j(A(\alpha^J), c)$ with respect to c evaluated at $c = 0$. Define the set

$$\mathcal{A} = \left\{ \alpha^J \in \mathbb{R}^J \mid g^{(l)}(A(\alpha^J), 0) \neq 0 \text{ for all integer } l \geq 1 \right\}. \quad (6)$$

This is the set of values of homogeneous terms, identified in the first stage, where the logit has nonzero derivatives and hence all moments of β are identified using Theorem 10. If $\mathcal{A} = \mathbb{R}^J$, then we would write the logit model is identified. Unfortunately, $\mathcal{A} \subset \mathbb{R}^J$, although we will show that $\mathbb{R}^J \setminus \mathcal{A}$ is a set of measure 0.

Assumption 14. *The support \mathcal{X} contains a nonempty open set of points (open in \mathbb{R}^K) of the form $(x'_1, \dots, x'_{j-1}, x'_j, x'_{j+1}, \dots, x'_J) = (0', \dots, 0', \tilde{x}'_j, 0', \dots, 0')$ surrounding $\tilde{x}_j = 0$.*

Theorem 15. *Let Assumptions 2, 7, 8, 11, and 14 hold. The homogeneous parameters $\alpha^{J,0}$ are identified. Then the true F^0 is identified for any $\alpha^J \in \mathcal{A}$. Further, \mathcal{A} is a set of measure 1 in \mathbb{R}^J .*

Identification of F^0 follows directly from Theorem 10 and the definition of \mathcal{A} once Assumption 8.1 is satisfied. A function that is real analytic is infinitely differentiable. The proof in the appendix shows that the set \mathcal{A} has measure 1. We note that α^J is always identified from the data at $x = 0$ and whether $\alpha^J \in \mathcal{A}^L$ can be computationally tested using computer algebra software, where

$$\mathcal{A}^L = \left\{ \alpha^J \in \mathbb{R}^J \mid g^{(l)}(A(\alpha^J), 0) \neq 0 \text{ for all integer } 1 \leq l \leq L \right\}$$

and L is the maximum order of the derivative considered by the computer. As variation in w causes variation in $\alpha_j = \alpha + w'_j \gamma_w$, variation in w (allowing regressors without random coefficients) ensures that the true F^0 is identified, not just identified for any $\alpha^J \in \mathcal{A}$.

As stated previously, our constructive identification argument uses only variation in product characteristics x_j around zero or around zero after centering, e.g., we can redefine $x_{j,k} = x_{j,k} - \bar{x}_k$ where \bar{x}_k is a constant term that is the same across the inside goods. Because we do not advocate analog estimation (instead preferring our mixtures estimator in Bajari, Fox, Kim, and Ryan (2010) or another mixtures estimator), we do not see using thin slices of data in identification as a problem for estimation.

6 Extensions

6.1 Identification for Higher Order Terms and Interactions

Our identification results do not extend to models with higher order polynomial terms and interaction terms. Our support conditions rule out these models for both the non-constructive and constructive identification arguments. When $x_{j,k}$ includes higher order terms or interactions of characteristics, the support of x_j cannot include a nonempty open set because $x_{j,k}$ has components that are nonlinearly dependent. This is a limitation of our identification results.

6.2 Need for a Special Regressor for a Random Coefficient for the Constant Term

Consider the utility specification $u_{i,j} = \alpha_i + x'_j \beta_i + \epsilon_{i,j}$, where we now allow for a random coefficient α_i on the constant term, which affects the utility of the inside relative to the outside goods. Consider the case of $J = 1$, or one inside and one outside good. In this case, the unspecified density $f_\alpha(\alpha)$ subsumes the type I extreme value density on $\epsilon_{i,j}$. This puts us in the model of Ichimura and Thompson (1997), where all previous identification results require special regressors (regressors with large support).

6.3 Endogeneity

If there is endogeneity in price due a demand shock or omitted product characteristic ξ_j in the utility of choice j , one can adopt Kim and Petrin (2009)'s control function approach. Kim and Petrin show that under a set of conditions (including restrictions on the supply side), one can include a proxy for ξ_j that is a (to be identified) nonparametric function of the residuals for all products from first stage regressions of pricing equations. Alternatively, one can identify ξ_j in a first stage using both a special regressor and instruments with the approach of Berry and Haile (2010). Either way, ξ_j or proxies for it can be added to the characteristic vector x_j and identification using our approach can be considered.

7 Conclusions

The random coefficients logit model has been used in empirical studies for over thirty years. In contrast to other work on identification in binary and multinomial choice, we exploit the type I extreme value distribution on the additive errors to show that the density of random coefficients is nonparametrically identified under an alternative set of assumptions, which are non-nested with those used in other approaches. By exploiting this special structure, we eliminate assumptions about the large support and the signs of coefficients on special regressors. We do use identification at one particular point ($x = 0$), but, in the non-constructive identification theorem, only to identify homogeneous parameters such as the product intercepts α in a first stage and to focus on variation in the characteristics of only one choice j . An open set around $x = 0$ is the only source of identification for our constructive identification result. From an econometric theory perspective, either approach allows complete proofs for the consistency of nonparametric estimators of the distribution of random coefficients. The proof of identification is also comforting to empirical researchers. Prior to our theorem, it was not known whether variation in x was sufficient to identify the distribution of the random coefficients in these models. One possibility was that the normality assumptions typically imposed on the distribution were crucial to identification: without restricting attention to a particular parametric functional form, two different distributions of random coefficients would be consistent with the data in the model, even with data on a continuum of x . We show that indeed the random coefficients logit model is nonparametrically identified, which provides a solid econometric foundation for its widespread use in empirical work. We can condition on discrete covariates, but, like the rest of the literature, we cannot point identify the distribution of random coefficients on discrete characteristics. Our identification results do not extend to models with higher order terms or interaction terms.

A Proofs

A.1 Proof of Lemma 5

We rephrase the proof of Theorem 3.8 in Stinchcombe and White (1998) in terms of our identification problem, otherwise our proof is essentially the same.

The forward direction of Lemma 5 holds because a stronger definition of identification implies a weaker definition. For the reverse direction of Lemma 5, assume to the contrary: the distribution of random coefficients is identified uniformly over \mathcal{B} with a nonempty open set $\mathcal{T}_0 \subset \mathcal{T}$ but there exists a nonempty open set $\tilde{\mathcal{T}} \subset \mathcal{T}$ and a compact set $\tilde{\mathcal{B}}$ where identification fails. The lack of identification means that there exist $F^0, F^1 \in \mathcal{F}(\tilde{\mathcal{B}})$ such that $F^0 \neq F^1$ but $P(x, F^0) = P(x, F^1)$ for all $x \in \tilde{\mathcal{T}}$. It follows that $\Delta(x) \equiv \int_{\tilde{\mathcal{B}}} g(x'\beta) d(F^0(\beta) - F^1(\beta)) = 0$ for all $x \in \tilde{\mathcal{T}}$.

Because $\tilde{\mathcal{B}}$ is compact and $g(\cdot)$ is real analytic, $\Delta(x)$ is itself a real analytic function. If a real analytic function equals to 0 on an open set, it equals 0 everywhere. If $\Delta(x) = 0$ everywhere, then $F^0 \in \mathcal{F}(\tilde{\mathcal{B}})$ cannot be identified on the set \mathcal{T}_0 , which gives a contradiction.

A.2 Proof of Theorem 10

First we introduce some notation for gradients of arbitrary order, which we need because $F(\beta)$ has a vector of K arguments, β . Let t be a vector of length T . For a function $h(t)$, we denote the $1 \times K^v$ block vector of v -th order derivatives as $\nabla^v h(t)$. $\nabla^v h(t)$ is defined recursively so that the k -th block of $\nabla^v h(t)$ is the $1 \times T$ vector $h_k^v(t) = \partial h_k^{v-1}(\theta) / \partial t'$, where h_k^{v-1} is the k -th element of $\nabla^{v-1} h(t)$. Using a Kronecker product \otimes , we can write $\nabla^v h(t) = \underbrace{\frac{\partial^v h(t)}{\partial t' \otimes \partial t' \otimes \dots \otimes \partial t'}}_{v \text{ Kronecker product of } \partial t'}$.

Take the derivatives with respect to the covariates x on both sides of $P(x, F) = \int g(\alpha, x'\beta) dF(\beta)$ and evaluate the derivatives at $x = 0$. By Assumption 8, for any $v = 1, 2, \dots$ and the chain rule repeatedly applied to the linear index $x'\beta$,

$$\begin{aligned} \nabla^v P(x, F)|_{x=0} &= \int g^{(v)}(x'\beta) \Big|_{x=0} \{\beta' \otimes \beta' \otimes \dots \otimes \beta'\} dF(\beta) \\ &= g^{(v)}(0) \int \{\beta' \otimes \beta' \otimes \dots \otimes \beta'\} dF(\beta). \end{aligned} \tag{7}$$

For each v there are K^v equations. Recall g is a known function. Therefore, as long as $g^{(v)}(0)$ is nonzero and finite for all $v = 1, 2, \dots$, we obtain the v -th moments of β for all $v \geq 1$. Now by Assumption 7, F satisfies the Carleman condition. Therefore, F^0 is identified since a probability measure satisfying the Carleman condition is uniquely determined by its moments.

A.3 Proof of Theorem 15

Identification arises from identifying all moments, as in Theorem 10. We wish to show that the set \mathcal{A} as defined in (6) has measure 1 in \mathbb{R}^J .

Let D_c be the derivative operator. We suppress $A_j(\alpha^J)$'s dependence on α^J and write $A_j = A_j(\alpha^J)$. For this purpose, we first obtain the derivatives of $g_j(\alpha^J, c)$ with respect to c ,

$$\begin{aligned} D_c g_j(\alpha^J, c) &= (A_j + e^c)^{-2} A_j e^c, \quad D_c^2 g_j(\alpha^J, c) = (A_j + e^c)^{-3} (A_j^2 e^c - A_j e^{2c}) \\ D_c^3 g_j(\alpha^J, c) &= (A_j + e^c)^{-4} (A_j^3 e^c - 4A_j^2 e^{2c} + A_j e^{3c}), \dots \end{aligned}$$

For $p \geq 3$, now we write the $(p-1)$ -th derivative as $D_c^{p-1} g_j(\alpha^J, c) = (A_j + e^c)^{-p} \sum_{l=1}^{p-1} \gamma_{p-l}^{(p)} A_j^{p-l} e^{lc}$. Then, we can write the p -th derivative as

$$D_c^p g_j(\alpha^J, c) = \left[\frac{1}{(A_j + e^c)^{p+1}} \sum_{l=1}^p \gamma_{p+1-l}^{(p+1)} A_j^{p+1-l} e^{lc} \right] \quad (8)$$

$$= D_c D_c^{p-1} g_j(\alpha^J, c)$$

$$= D_c \left[\frac{1}{(A_j + e^c)^p} \sum_{l=1}^{p-1} \gamma_{p-l}^{(p)} A_j^{p-l} e^{lc} \right]$$

$$= \frac{1}{(A_j + e^c)^p} \sum_{l=1}^{p-1} l \gamma_{p-l}^{(p)} A_j^{p-l} e^{lc} - \frac{1}{(A_j + e^c)^{p+1}} p \sum_{l=1}^{p-1} \gamma_{p-l}^{(p)} A_j^{p-l} e^{(l+1)c}$$

$$= \frac{1}{(A_j + e^c)^{p+1}} \left((A_j + e^c) \sum_{l=1}^{p-1} l \gamma_{p-l}^{(p)} A_j^{p-l} e^{lc} - \sum_{l=1}^{p-1} p \gamma_{p-l}^{(p)} A_j^{p-l} e^{(l+1)c} \right)$$

$$= \frac{1}{(A_j + e^c)^{p+1}} \left(\sum_{l=1}^{p-1} l \gamma_{p-l}^{(p)} A_j^{p+1-l} e^{lc} + \sum_{l=1}^{p-1} l \gamma_{p-l}^{(p)} A_j^{p-l} e^{(l+1)c} - \sum_{l=1}^{p-1} p \gamma_{p-l}^{(p)} A_j^{p-l} e^{(l+1)c} \right)$$

$$= \frac{1}{(A_j + e^c)^{p+1}} \left(\begin{aligned} &\gamma_{p-1}^{(p)} A_j^p e^c + \sum_{l'=2}^{p-1} l' \gamma_{p-l'}^{(p)} A_j^{p+1-l'} e^{l'c} \\ &+ \sum_{l=1}^{p-1} l \gamma_{p-l}^{(p)} A_j^{p-l} e^{(l+1)c} - \sum_{l=1}^{p-1} p \gamma_{p-l}^{(p)} A_j^{p-l} e^{(l+1)c} \end{aligned} \right) \quad (9)$$

$$= \frac{1}{(A_j + e^c)^{p+1}} \left(\begin{aligned} &\gamma_{p-1}^{(p)} A_j^p e^c + \sum_{l=1}^{p-2} (l+1) \gamma_{p-l-1}^{(p)} A_j^{p-l} e^{(l+1)c} \\ &+ \sum_{l=1}^{p-1} l \gamma_{p-l}^{(p)} A_j^{p-l} e^{(l+1)c} - \sum_{l=1}^{p-1} p \gamma_{p-l}^{(p)} A_j^{p-l} e^{(l+1)c} \end{aligned} \right) \quad (10)$$

$$= \frac{1}{(A_j + e^c)^{p+1}} \left(\begin{aligned} &\gamma_{p-1}^{(p)} A_j^p e^c - \gamma_1^{(p)} A_j^1 e^{pc} \\ &+ \sum_{l=1}^{p-2} \left\{ (l+1) \gamma_{p-l-1}^{(p)} + l \gamma_{p-l}^{(p)} - p \gamma_{p-l}^{(p)} \right\} A_j^{p-l} e^{(l+1)c} \end{aligned} \right), \quad (11)$$

where in (9) and (10), we take out the first element in the first sum and change the index l' to $l+1$. (11) is obtained by rearranging terms and collecting coefficients on $A_j^{p-l} e^{(l+1)c}$ for $j = 1$ to $p-2$.

To fix the undetermined coefficients $\gamma_{p-l}^{(p)}$'s, we compare the coefficients from (8) and (11) and

obtain

$$\begin{aligned}
\sum_{l=1}^p \gamma_{p+1-l}^{(p+1)} A_j^{p+1-l} e^{lc} &= \gamma_p^{(p+1)} A_j^p e^c + \sum_{l=2}^{p-1} \gamma_{p+1-l}^{(p+1)} A_j^{p+1-l} e^{lc} + \gamma_1^{(p+1)} A_j^1 e^{pc} \\
&= \gamma_p^{(p+1)} A_j^p e^c + \sum_{l=1}^{p-2} \gamma_{p-l}^{(p+1)} A_j^{p-l} e^{(l+1)c} + \gamma_1^{(p+1)} A_j^1 e^{pc} \\
&= \gamma_{p-1}^{(p)} A_j^p e^c + \sum_{l=1}^{p-2} \left\{ (l+1) \gamma_{p-l-1}^{(p)} + l \gamma_{p-l}^{(p)} - p \gamma_{p-l}^{(p)} \right\} A_j^{p-l} e^{(l+1)c} - \gamma_1^{(p)} A_j^1 e^{pc}.
\end{aligned}$$

We find

$$\gamma_p^{(p+1)} = \gamma_{p-1}^{(p)} \quad (12)$$

$$\gamma_{p-l}^{(p+1)} = (l+1) \gamma_{p-l-1}^{(p)} - (p-l) \gamma_{p-l}^{(p)} \text{ for } p \geq 3 \quad (13)$$

$$\gamma_1^{(p+1)} = -\gamma_1^{(p)}. \quad (14)$$

This system generates the coefficients for all $p \geq 1$. For the initial value, we obtain $\gamma_1^{(2)} = 1$. When $p = 2$, we find

$$\gamma_2^{(3)} = \gamma_1^{(2)} = 1, \gamma_1^{(3)} = -\gamma_1^{(2)} = -1$$

and when $p = 3$, we find

$$\gamma_3^{(4)} = \gamma_2^{(3)} = 1, \gamma_2^{(4)} = 2\gamma_1^{(3)} - 2\gamma_2^{(3)} = -4, \gamma_1^{(4)} = -\gamma_1^{(3)} = 1.$$

Now we examine whether $D_c^p g_j(\alpha^J, c)|_{c=0}$ can take the value of zero for some A_j (and hence for some α) at some p . For this purpose, we evaluate the derivatives at $c = 0$ ($x_j = 0$) and obtain expressions with respect to A_j for the p -th order derivative as $D_c^p g_j(\alpha^J, c)|_{c=0} = \frac{1}{(A_j+1)^{p+1}} \sum_{l=1}^p \gamma_{p+1-l}^{(p+1)} A_j^{p+1-l} = 0$ for all $p \geq 1$. This is equivalent to solving

$$\sum_{l=1}^p \gamma_{p+1-l}^{(p+1)} A_j^{p-l} = 0. \quad (15)$$

We consider two cases. In the first case where $\alpha_j = 0$ for all j , i.e., all goods are symmetric, the condition of nonzero derivatives is obtained using the rational zero test. In this case we have $A_j = J$ and note that the coefficient on A_j^{p-1} (the highest order term in the equation) in (15) is equal to $\gamma_p^{(p+1)} = 1$ for all p . Also note that the constant term (the coefficient on A_j^0 in (15)), $\gamma_1^{(p+1)}$, is equal to 1 when p is odd and is equal to -1 when p is even. By the rational zero test, this implies that the only possible positive rational number solution in (15) is $J = 1$. A positive integer greater than 1 cannot be the solution of (15) for any p . This proves the condition of nonzero derivatives for the symmetric case.

Second we consider the asymmetric case where α_j may not be 0. In this case we allow for

non-integer solutions of the equation (15).¹⁶ Nonetheless we will show that the set

$$\mathcal{A}^C = \{\alpha^J \mid D_c^p g_j(\alpha^J, c)|_{c=0} = 0 \text{ for at least one } p \geq 1, \alpha^J \in \mathbb{R}^J\}$$

has measure zero in \mathbb{R}^J .

Note that the set of values $\mathbb{A} \subset \mathbb{R}_+$ that collect values of A_j that solve the equation (15) for at least one p is countable because for any order p the equation (15) has at most p number of solutions, i.e., we can find an injective function that maps \mathbb{A} to \mathbb{N} . Now consider any element $\tilde{A} \in \mathbb{A}$. We claim that the set of values $\tilde{\mathcal{A}}(\tilde{A}) = \{\alpha^J \mid A_j(\alpha^J) = \tilde{A}, \alpha^J \in \mathbb{R}^J\}$ has measure zero in \mathbb{R}^J because $\tilde{\mathcal{A}}$ is at most a subset of the vector space of \mathbb{R}^{J-1} . By construction, we have $\mathcal{A}^C = \cup_{\tilde{A} \in \mathbb{A}} \tilde{\mathcal{A}}(\tilde{A}) = \cup_{\tilde{A} \in \mathbb{A}} \{\alpha^J \mid A_j(\alpha^J) = \tilde{A}, \alpha^J \in \mathbb{R}^J\}$. Finally we conclude that the set \mathcal{A}^C has measure zero in \mathbb{R}^J because a countable union of measure zero sets has measure zero, as $\tilde{\mathcal{A}}(\tilde{A})$ has measure zero for all $\tilde{A} \in \mathbb{A}$ and the set \mathbb{A} is countable. This completes the proof.

¹⁶Because the equation (15) holds for any j , in practice we may pick j such that $j^* = \operatorname{argmin}\{\alpha_1, \dots, \alpha_J\}$. Then $A_{j^*} = \exp(-\alpha_{j^*}) + \sum_{j' \neq j^*} \exp(\alpha_{j'} - \alpha_{j^*}) > J - 1$. This again rules out all rational roots of (15) with $J \geq 2$ because the only rational root of (15) is $A_{j^*} = 1$ according to the rational zero test.

If A_{j^*} is not a solution of (15) for any order of derivative, we satisfy the nonzero derivatives condition in Assumption 8 by focusing on product j^* . Therefore, suppose A_{j^*} is a solution to the equation (15) corresponding to the p -th order derivative. Then we cannot identify the p -th moments of $F^0(\beta)$ using our strategy because $D_c^p g_{j^*}(c=0) = 0$. But we can also utilize the share equation of other goods (including the outside good) to identify the p -th moments of $f(\beta)$. We can choose a j' -th good such that (i) $A_{j'} = \exp(-\alpha_{j'}) + \sum_{j \neq j'} \exp(\alpha_j - \alpha_{j'}) \neq A_{j^*}$ and (ii) $A_{j'}$ does not solve the equation (15) corresponding to the p -th order derivative. Then we can identify the p -th order moments of $F^0(\beta)$ from the share equation of the j' -th good. Therefore, combining these two share equations, we can identify all the moments of $F^0(\beta)$.

Moreover, we conjecture (we have experimented with this) that if A_{j^*} solves (15) for the p -th order derivative, A_{j^*} cannot be a solution to (15) for any other order derivative than the p -th order. Therefore all the moments of $F^0(\beta)$ other than the p -th order moments are identified using derivatives of the share equation for the j^* -th good.

References

- [1] Andrews, D.W.K and Schafgans, M.M.A. (1998), “Semiparametric estimation of the intercept of a sample selection model”, *Review of Economic Studies*, 65(3), 497–517.
- [2] Bajari, P., J. Fox, K. Kim, and S. Ryan (2010), “A Simple Estimator for the Distribution of Random Coefficients”, University of Minnesota working paper.
- [3] Berry, S and Haile, P. (2010), “Nonparametric Identification of Multinomial Choice Models with Heterogeneous Consumers and Endogeneity”, Yale University working paper.
- [4] Boyd, JH and RE Mellman (1980), “Effect of Fuel Economy Standards on the U. S. Automotive Market: An Hedonic Demand Analysis”, *Transportation Research B*, 14(5), 367–378.
- [5] Briesch, Richard A., Pradeep K. Chintagunta, and Rosa L. Matzkin (2010), “Nonparametric Discrete Choice Models with Unobserved Heterogeneity”, *Journal of Business and Economic Statistics*, 28(2), 291–307.
- [6] Burda, Martin, Matthew Harding and Jerry Hausman (2008), “A Bayesian Mixed Logit-Probit for Multinomial Choice”, *Journal of Econometrics*, 147(2), 232–246.
- [7] Cardell, NS and FC Dunbar (1980), “Measuring the societal impacts of automobile downsizing”, *Transportation Research B*, 14(5), 423–434.
- [8] Cramér, H. and H. Wold (1936), “Some Theorems on Distribution Functions”, *Journal of the London Mathematical Society*, s1-11(4), 290–294.
- [9] Chiappori, PA and I. Komunjer (2009), “On the Nonparametric Identification of Multiple Choice Models”, Columbia University working paper.
- [10] Domencich, Thomas A. and Daniel L. McFadden (1975), *Urban travel demand : a behavioral analysis : a Charles River Associates research study*. North Holland.
- [11] Fox, Jeremy T. and Amit Gandhi (2009), “Nonparametric Identification and Estimation of Random Coefficients in Nonlinear Economic Models”, University of Michigan working paper.
- [12] Gautier, Eric and Kitamura, Yuichi (2009), “Nonparametric Estimation in Random Coefficients Binary Choice Models”, Cowles Foundation Discussion Paper No. 1721.
- [13] Hausman, Jerry and David A. Wise (1978), “A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences”, *Econometrica*, 46(2), 403–426.
- [14] Heckman, JJ and RJ Willis (1977), “A beta-logistic model for the analysis of sequential labor force participation by married women”, *Journal of Political Economy*, 85(1), 27–58.

- [15] Hornik, Kurt (1991), “Approximation capabilities of multilayer feedforward networks”, *Neural Networks*, 4, 251–257.
- [16] Ichimura, Hidehiko and T. Scott Thompson (1998), “Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution,” *Journal of Econometrics*, 86(2), 269–295.
- [17] Kahn, S. and E. Tamer (2010), “Irregular Identification, Support Conditions, and Inverse Weight Estimation,” forthcoming in *Econometrica*.
- [18] Kim, Kyoo il and Amil Petrin (2009), “Control Function Corrections for Unobserved Factors in Differentiated Product Models”, University of Minnesota working paper.
- [19] Krantz, Steven G. and Harold R. Parks (2002), *A Primer on Real Analytic Functions*, Second Edition. Birkhauser.
- [20] Lewbel, Arthur. (2000) “Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables”, *Journal of Econometrics*, 97, 1, 145–177.
- [21] Magnac, T and E. Maurin (2007), “Identification and information in monotone binary models”, *Journal of Econometrics*, 139(1), 76–104.
- [22] Matzkin, Rosa L. (2007), “Heterogeneous Choice”, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress - Volume III*, Ch. 4.
- [23] McFadden, Daniel and Kenneth Train (2000), “Mixed MNL models for discrete response”, *Journal of Applied Econometrics*, 15(5): 447–470.
- [24] Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2005), *Bayesian Statistics and Marketing*. West Sussex: John Willy & Sons.
- [25] Shohat, J.A. and Tamarkin, J.D. (1943), *The Problem of Moments*, American Mathematics Society, Providence, RI.
- [26] Stinchcombe, Maxwell B and Halbert White, (1998), “Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative”, *Econometric Theory*, 14, 295–325.
- [27] Teicher, Henry (1963), “Identifiability of Finite Mixtures”, *Annals of Mathematical Statistics*, 34, 1265–1269.
- [28] Train, K.E. (2008), “EM Algorithms for Nonparametric Estimation of Mixing Distributions”, *Journal of Choice Modelling*, 1(1), 40–69.