Identification of the Distribution of Random Coefficients in Static and Dynamic Discrete Choice Models

Kyoo il Kim*
University of Minnesota

May 2014

Abstract

We show that the distributions of random coefficients in various discrete choice models are nonparametrically identified. Our identification results apply to static discrete choice models including binary logit, multinomial logit, nested logit, and probit models as well as dynamic programming discrete choice models. In these models the only key condition we need to verify for identification is that the type specific model choice probability belongs to a class of functions that include analytic functions. Therefore our identification results are general enough to include most of commonly used discrete choice models in the literature. Our identification argument builds on insights from nonparametric specification testing. We find that the role of analytic function in our identification results is to effectively remove the full support requirement often exploited in other identification approaches, which is very important for discrete choice models where the values of covariates are often bounded below and above.

Keywords: Random Coefficients, Nonparametric Identification, Logit and Probit, Discrete Choice, Dynamic Discrete Choice

*Corresponding author: Kyoo il Kim, Department of Economics, University of Minnesota, 4-129 Hanson Hall, 1925 Fourth Street South, Minneapolis, MN 55455, Tel: +1 612 625 6793, E-mail: kyookim@umn.edu.

1 Introduction

Modeling heterogeneity in preferences of economic agents has been of significant interests in both theoretical and empirical studies where otherwise identical agents behave differently when faced with identical choice environments. A growing econometric literature has addressed this problem by providing estimators that allow the coefficients of the economic model to vary across agents. Random coefficients have been popularly used to address this individual heterogeneity. For recent work in discrete choice estimations with random coefficients including consumer demands, see (e.g.) Berry, Levinsohn, and Pakes (1995), Nevo (2001), Petrin (2002), Rossi, Allenby, and McCulloch (2005), Lewbel (2000), Burda, Harding, and Hausman (2008), McFadden and Train (2000), Briesch, Chintagunta, and Matzkin (2010), Hoderlein, Klemela, and Mammen (2010), and Gautier and Kitamura (2013). However, identification studies on random coefficient models, which can be applied to various discrete choice models - as commonly used in empirical studies - still have been scarce with only a few exceptions. Moreover, there has been no unifying identification framework that can be generally applied to a variety of discrete choice models. In this paper we provide one such important result.

Building on insights from nonparametric specification testing literature (e.g. Stinchcombe and White 1998, Bierens 1982, 1990) we show that the distributions of random coefficients in discrete choice models are nonparametrically identified if the type specific choice probability satisfies the property that the span of the type specific choice probabilities is weakly dense in the space of bounded and continuous functions. We then show that this identification condition is satisfied under three conditions. The first is that the type specific model choice probability is a real analytic function and the support of the distribution of covariates (e.g. characteristics of products) is a nonempty open set. The second is that the function inside the type specific choice probability is monotonic in each element of the covariates vector that has random coefficients. This condition is trivially satisfied for static discrete choice models with index restrictions. Importantly we verify this monotonicity condition also holds for dynamic discrete choice models. Therefore, the second condition is not restrictive for most of discrete choice models that are commonly used in the literature. The third and last condition is that we need at least one value of covariates such that the type specific choice probability does not depend on random coefficients at this particular value of covariates. To satisfy this condition we can typically let the covariates include the value of zero or re-center the covariates at zero. We find these three identifying conditions are satisfied for a class of discrete choice logit models including binary choice, multinomial choice, nested logit, and dynamic programming discrete choice models. The required condition of being a real analytic function is sufficient but not necessary. As an example we find that the distribution of random coefficients in the probit model is also nonparametrically identified but the probit function is not analytic.

Our identification argument differs from the "identification at infinity" using a special covariate (e.g. Lewbel 2000) and from the Cramer-Wold device (e.g. Ichimura and Thompson 1997). Berry and Haile (2010) also provide an important identification result for discrete choice models but they require a special covariate and its full support condition while they do not use the logit structure. Moreover, their identification objects of interest do not include the distribution of random coefficients. In our opinion, the main concern with the special regressor is the requirement for large support. Large supports are sometimes acceptable but not often supported in typical datasets used in discrete choice estimation. Our study focuses on the nonparametric identification of distribution of random coefficients while our identification strategy explicitly resorts to the logit or the probit error structure, as is common in empirical work.[1] This 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. The framework we use is similar to Fox, Kim, Ryan, and Bajari (2012) but their results are only specific to the static multinomial logit model and our framework extends to other discrete choice models including nested logit, probit, and dynamic discrete choices. To our knowledge, our work is the first paper to formally show the nonparametric identification of random coefficients in dynamic programming discrete choice models. Our identification results are general enough to include most of commonly used discrete choice models and also can be used to develop a sieve approximation based estimator of the nonparametric distribution as in Fox, Kim, and Yang (2013). Although our identification results are not constructive, the results can be used to verify identification conditions for the consistency of the sieve approximation based estimator in Fox, Kim, and Yang (2013).

The organization of the paper is as follows. In Section 2 we review various discrete models that fit into our identification framework. Section 3 develops the identification theorems. In Section 4 we show that the identification conditions are satisfied for various static discrete choice models. In Section 5 we show the identification conditions hold for the dynamic discrete choice model. In Section 6 we conclude. Technical details are gathered in the Appendix.

2 Discrete Choice Models with Random Coefficients

Here we review various examples of discrete choice models with random coefficients to which our identification theorems in Section 3 are applied. These models are mostly commonly used in empirical studies.

2.1 Logit model with individual choices

The motivating example is the multinomial logit discrete choice - including binary choice - with random coefficients where agents $i = 1, \ldots, N$ can choose between $j = 1, \ldots, J$ mutually

[1] Other important identification studies in static discrete choice models include Briesch, Chintagunta, and Matzkin (2010), Chiappori and Komunjer (2009), Gautier and Kitamura (2013), and Fox and Gandhi (2010) but their modeling primitives are all different from our focus in this paper.

exclusive alternatives and one outside option (e.g. outside good). The random coefficients logit model was first proposed by Boyd and Mellman (1980) and Cardell and Dunbar (1980).

In the random coefficients model, the preference parameter β_i is distributed by $F(\beta)$ and is independent of the exogenous covariates. The exogenous covariates for choice j are in the $K \times 1$ vector $x_{i,j}$. We let $x_i = (x'_{i,1}, \ldots, x'_{i,J})$. This distribution $F(\beta)$ is the object of our interest. In the random utility model, agent i of type β_i has her utility of choosing alternative j is equal to

$$u_{i,j} = \alpha + x'_{i,j}\beta_i + \epsilon_{i,j} \tag{1}$$

where α is the non-random constant term, so this model does not allow random coefficient for the constant term. Assume that $\epsilon_{i,j}$ is distributed as Type I extreme value including an outside good with utility $u_{i.0} = \epsilon_{i,0}$ and agents are the utility maximizers. Then the outcome variable $y_{i,j}$ is defined as

$$y_{i,j} = \begin{cases} 1 & \text{if } u_{i,j} > u_{i,j'} \text{ for all } j' \neq j \\ 0 & \text{otherwise} \end{cases}.$$

The type specific choice probability of taking choice j at x_i is

$$g_j(x_i, \beta, \alpha) = \frac{\exp(\alpha + x'_{i,j}\beta)}{1 + \sum_{j'=1}^{J} \exp(\alpha + x'_{i,j'}\beta)}.$$

In the data we observe the conditional choice probability of the mixture $P(y_{i,j} = 1|x_i)$ and the logit model implies that

$$P(y_{i,j} = 1|x_i) = \int g_j(x_i, \beta, \alpha) dF(\beta) = \int \frac{\exp(\alpha + x'_{i,j}\beta)}{1 + \sum_{j'=1}^{J} \exp(\alpha + x'_{i,j'}\beta)} dF(\beta). \tag{2}$$

Our key question is whether we can identify $F(\beta)$ from the observed $P(y_{i,j} = 1|x_i)$ and the type specific model choice probability $g_j(x_i, \beta, \alpha)$ in (2). In Section 4.1 we show $F(\beta)$ is identified for this multinomial logit model.

2.2 Nested logit model with individual choices

We consider a nested logit model with the following random utility

$$u_{i,j,l} = z'_{i,j}\gamma_i + x'_{i,j,l}\beta_{j,i} + \epsilon_{i,j,l}$$

for $l = 1, \ldots, L_j$ choices per group j with $j = 1, \ldots, J$ groups of choices and $j = 0$ being the outside good ($u_{i,0} = \epsilon_{i,0}$) where $z_{i,j}$ denotes the group specific covariates while $x_{i,j,l}$ denotes the choice specific covariates. Let $z_i = (z'_{i,1}, \ldots, z'_{i,J})$ and $x_i = (x'_{i,1,1}, \ldots, x'_{i,1,L_1}, \ldots, x'_{i,J,L_J})$.

The nested logit model allows individual tastes to be correlated across products in each group. The error terms follow a generalized extreme value distribution (McFadden 1978) of the form

$$F(\varepsilon)=\exp\left(-\sum_{j=0}^{J}\left(\sum_{l=1}^{L_j}\exp(-\varepsilon_{j,l}/\rho_j)\right)^{\rho_j}\right)$$

where ρ_j reflects the correlation between $\varepsilon_{j,l}$ and $\varepsilon_{j,l'}$ as $\rho_j=\sqrt{1-Corr[\varepsilon_{j,l},\varepsilon_{j,l'}]}$ for all $l\neq l'$and $(L_0=1,\rho_0=1)$ for the outside good.

The type specific choice probability of taking choice l in the j category at (z_i,x_i) is

$$g_{j,l}(z_i,x_i,\gamma,\beta,\rho)=\frac{\exp\left(z'_{i,j}\gamma+\rho_j\log\left(\sum_{l'=1}^{L_j}\exp(x'_{i,j,l'}\beta_j/\rho_j)\right)\right)}{\sum_{j'=0}^{J}\exp\left(z'_{i,j'}\gamma+\rho_{j'}\log\left(\sum_{l'=1}^{L_{j'}}\exp(x'_{i,j',l'}\beta_{j'}/\rho_{j'})\right)\right)}\frac{\exp(x'_{i,j,l}\beta_j/\rho_j)}{\sum_{l'=1}^{L_j}\exp(x'_{i,j,l'}\beta_j/\rho_j)}$$

where $\beta=(\beta'_1,\ldots,\beta'_J)$ and $\rho=(\rho_1,\ldots,\rho_J)$. We have

$$P(y_{i,j,l}=1|z_i,x_i)=\int\cdots\int g_{j,l}(z_i,x_i,\gamma,\beta,\rho)dF_\gamma(\gamma)dF_{\beta_1}(\beta_1)\cdots dF_{\beta_J}(\beta_J) \tag{3}$$

where $F_\gamma(\gamma)$ and $F_{\beta_j}(\beta_j)$'s are distribution functions of γ and β_j's, so we assume γ and β_j's are independent each other while we allow distributions of components inside β_j's can be dependent. In Section 4.2 we show that the distribution of random coefficients $F_\gamma(\gamma)$ and $F_{\beta_j}(\beta_j)$'s are identified for this nested logit model.

2.3 Probit model with binary choice

When $\epsilon_{i,j}$ in (1) follows a standard normal distribution with $J=1$ and $u_{i.0}=0$. The model becomes a probit binary choice. We have

$$P(y_{i,1}=1|x_{i,1})=\int\Phi(\alpha+x'_{i,1}\beta)dF(\beta)$$

where $\Phi(\cdot)$ denotes the CDF of standard normal. In Section 4.3 we show this probit model is also identified.

2.4 Logit model with aggregate data

The multinomial logit model can be used when data only on market shares s_j's are available but individual level data are not. We assume the utility of agent i is

$$u_{i,j}=\alpha+x'_j\beta_i+\epsilon_{i,j}$$

where β is distributed by $F(\beta)$. In this case the logit model implies

$$s_j=\int g_j(x,\beta,\alpha)dF(\beta)=\int\frac{\exp(\alpha+x'_j\beta)}{1+\sum_{j'=1}^{J}\exp(\alpha+x'_{i,j'}\beta)}dF(\beta).$$

2.5 Dynamic discrete choice models

We consider the identification of distribution of random coefficients in dynamic discrete choice models (e.g. Rust 1987, 1994) - note that the original models of Rust do not have random coefficients. We assume that per period utility of agent i in a period t from choosing action $d \in D$ is

$$u_{i,d,t} = x'_{i,d,t}\theta + \epsilon_{i,d,t}.$$

The error term is iid extreme value across agents, choices, and time periods and possibly a subset of θ, β is distributed as $F(\beta)$. We let $x_{i,t} = (x'_{i,1,t}, \ldots, x'_{i,|D|,t})$. The type specific conditional choice probability is

$$g_d(x_{i,t}, \theta) = \frac{\exp(v(d, x_{i,t}, \theta))}{\sum_{d'=1}^{|D|} \exp(v(d', x_{i,t}, \theta))} \tag{4}$$

where $v(d, x_{i,t}, \theta)$ denotes Rust's choice-specific value function.

As an illustration consider a dynamic binary choice model of Rust (1987) where the conditional choice probability of taking an action "1" is given by

$$g_1(x, \beta, \alpha) = \frac{\exp\{x'\beta + \delta EV(x, 1; \beta, \alpha)\}}{\exp\{\alpha + \delta EV(x, 0; \beta, \alpha)\} + \exp\{x'\beta + \delta EV(x, 1; \beta, \alpha)\}} \tag{5}$$

$$= \frac{\exp\{x'\beta + \delta\left[EV(x, 1; \beta, \alpha) - EV(x, 0; \beta, \alpha)\right]\}}{\exp\{\alpha\} + \exp\{x'\beta + \delta\left[EV(x, 1; \beta, \alpha) - EV(x, 0; \beta, \alpha)\right]\}} \tag{6}$$

and $EV(x, d; \beta, \alpha)$ is given by the unique solution to the Bellman equation

$$EV(x, d; \beta, \alpha) = \int_y \log\left\{\exp\{y'\beta + \delta EV(y, 1; \beta, \alpha)\} + \exp\{\alpha + \delta EV(y, 0; \beta, \alpha)\}\right\} \pi(dy|x, d) \tag{7}$$

with the transition density $\pi(dy|x, d)$. Note that although the distribution of β does not depend on x, the evolution of the state variable x over time depends on the type specific value β. This is because individuals having the same $x_{i,t} = \widetilde{x}$ at time t but having different β's will make different choices at time t and their states in the following time periods will be different. However, we note that in the evaluation of value functions in (7), we need only the transition density of states and this transition density is independent of β given D because β affects the transition of states only through the choice d. Therefore, the transition density $\pi(dy|x, d)$ is not a function of β, which is typically identified in a pre-stage of estimation.

In Rust (1987)'s bus engine replacement example, $d = 0$ denotes the replacement of an engine, α denotes the scrap value, and β is the unit operation cost with mileage equal to x. When the random coefficient β is distributed with $F(\beta)$, we have

$$P(1|x) = \int g_1(x, \beta, \alpha) dF(\beta) \tag{8}$$

where $P(1|x)$ is the true (population) conditional choice probability.

We study identification of these dynamic discrete choice models with random coefficients in Section 5.

3 Identification

In a general framework we develop nonparametric identification of the distribution of random coefficients $F(\beta)$ in discrete choice models. We then apply the results to the models of Section 2. The econometrician observes covariates or characteristics x and the probability of some discrete outcome indicators y, denoted by $G(x)$. Since our leading example is discrete choice models, we interpret $G(x)$ as the conditional choice probability and let $g(x,\beta)$ be the probability of an agent with the random coefficient β taking the choice. We assume that β and x are independent.

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

$$G(x) = \int h(x,\beta)\, dF(\beta) \tag{9}$$

where $h(x,\beta)$ is a known function of (x,β). Identification means a unique $F(\beta)$ solves this equation for all x. Let $G_0(x)$ denote the true function of $G(x)$ and let $F_0(\beta)$ denote the true function of $F(\beta)$ such that

$$G_0(x) = \int h(x,\beta)\, dF_0(\beta).$$

Then the identification means for any $F_1 \neq F_0$, we must have $G_1 \neq G_0$. Because $G_0(x) = E[y|x]$ is nonparametrically identified, we focus on the identification of F_0 below treating G_0 is known.

To formalize the notion of identification we develop notation as follows. First let ρ be any metric on the space of finite measures inducing the weak convergence of measures. For example, this includes the Lévy-Prokhorov metric for distribution functions. Further define

$$\mathcal{H} = \left\{ h(x,\beta) : R^{2K} \to R \,:\, x \in \mathcal{X} \subset \mathbb{R}^K, \beta \in \mathcal{B} \subset \mathbb{R}^K \right\}.$$

Note that $h \in \mathcal{H}$ is read as a function of β given x and is also a function of x given β. Then the identification means $F_1 = F_0$ in the weak topology if and only if $\int h dF_1 = \int h dF_0$ for all $h \in \mathcal{H}$. Let $C(\mathcal{B})$ be the set of continuous and bounded functions on $\mathcal{B}$, the support of $F(\beta)$. We let $\mathcal{F}(\mathcal{B})$ be the set of continuous and bounded distribution functions, supported on $\mathcal{B}$. We further let $\mathcal{G}(\mathcal{X})$ be the space of continuous and bounded functions on $\mathcal{X}$, generated by the mixture of (9) and assume every $G \in \mathcal{G}(\mathcal{X})$ is measurable with a measure μ. We let $\mathcal{F}(\mathcal{B})$ be endowed with the metric $\rho(F_0, F_1)$ for $F_0, F_1 \in \mathcal{F}(\mathcal{B})$ and $\mathcal{G}(\mathcal{X})$ be endowed with the metric $d(G_1, G_2)$ for $G_1, G_2 \in \mathcal{G}(\mathcal{X})$. We also assume that every $h \in \mathcal{H}$ is measurable with respect to $F \in \mathcal{F}(\mathcal{B})$ for almost every $x \in \mathcal{X}$. Finally let $\text{sp}\mathcal{H}$ denote the span of $\mathcal{H}$.

Now suppose $\mathcal{H}$ satisfies that for all $h \in C(\mathcal{B})$, for all $F \in \mathcal{F}(\mathcal{B})$, and for all $\delta > 0$, we can find a $h' \in \mathrm{sp}\mathcal{H}$ such that

$$\left| \int h' dF - \int h dF \right| < \delta. \tag{10}$$

Then by the definition of the span and the linearity of the integral, the condition (10) implies that $F_1 = F_0$ (in the weak topology) if and only if $\int h dF_1 = \int h dF_0$ for all $h \in \mathcal{H}$. This is because the condition $\int h dF_1 = \int h dF_0$ for all $h \in \mathcal{H}$ becomes equivalent to $\int h dF_1 = \int h dF_0$ for all $h \in C(\mathcal{B})$ under (10). This means that our identification condition is equivalent to the condition that the linear span of $\mathcal{H}$ is weakly dense in $C(\mathcal{B})$. We will show that some class of functions of $h(x, \beta)$ satisfy this weak denseness. We then show the type specific model choice probabilities of various discrete choice models - commonly used in empirical studies - belong to this class. Therefore, characterizing the class of functions $h(x, \beta)$ that satisfy the weak denseness becomes our tool for identification of the distribution of random coefficients.

3.1 Identification with known support of the distribution

First we consider the identification problem when the support of the distribution of the random coefficients $\mathcal{B}$ is known. Then we relax this arguably strong assumption in Section 3.3. We define our notion of identification formally.

Definition 1. For given $F \neq F_0$, $h \in \mathcal{H}$ distinguishes F if $d(G, G_0) \neq 0$. If for any $F(\neq F_0) \in \mathcal{F}$ there exists a distinguishing $h \in \mathcal{H}$, then $\mathcal{H}$ is totally distinguishing. If for any $F(\neq F_0) \in \mathcal{F}$, all but a negligible set of $h \in \mathcal{H}$ are distinguishing, then $\mathcal{H}$ is generically totally distinguishing.

The implication of $\mathcal{H}$ being generically totally distinguishing is that then F_0 is identified on any subset $\tilde{\mathcal{X}} \subset \mathcal{X}$ with $\mu(\tilde{\mathcal{X}}) \neq 0$.

We note that this notion of identification is closely related to the notion of *revealing* and totally *revealing* in the consistent specification testing problem of Stinchcombe and White (1998) and works of Bierens (1982, 1990). We first lay out our identification theorem below (Theorem 1) and note that its proof is closely related with Theorem 2.3 in Stinchcombe and White (1998) since the class of $\mathcal{H}$ that is generically totally revealing in Stinchcombe and White (1998) is generically totally distinguishing in our problem of identification.

But there are several important differences need to be pointed out. First their problem is a consistent specification testing where *the index set* $\mathcal{B}$ and draw of β's (not necessarily random) are arbitrary choices of a researcher, so the distribution of β is not of their interest at all but our problem is the identification of the distribution of β. Second we switch the role of x and β in the specification testing problems such that x's in $\mathcal{X}$ now generate the functions in $\mathcal{H}$. The last key difference is that for our identification result we do not need to restrict the function $h(x, \beta)$ to take the form of $h(x, \beta) = g(x_1 + \tilde{x}'\beta)$ (i.e., affine in β). This requires a normalization of coefficient for (e.g.) a *special regressor* x_1. This will be replaced by the requirement that $\mathcal{X}$ includes at least one value x^* such that $h(x, \beta)$ does not depend on the random coefficients β at x^* in our identification. Note that without loss of generality, following our leading example of the logit models, we can take $x^* = 0$ or re-center x at zero such that $\{0\} \subset \mathcal{X}$ for linear index models of the form $h(x, \beta) = g(x'\beta)$. Our first identification theorem becomes

Theorem 1. Let $\mathcal{H}_g = \{h : h(x, \beta) = g(x'\beta), x \in \mathcal{X}\}$ where (i) $\mathcal{X} \subset \mathbb{R}^K$ is a nonempty open set, (ii) $\{0\} \subset \mathcal{X}$, and (iii) g is real analytic. Suppose $\mathcal{B}$ is known. Then $\mathcal{H}_g$ is generically totally distinguishing if and only if g is non-polynomial. Moreover, $\mathcal{H}_g$ is also totally distinguishing.

Proof. Theorem 1 is implied by Theorem 3 below and hence the proof is omitted. We prove only Theorem 3 in Section 3.4. □

In the theorem we restrict our attention to the class of models with the linear index inside the model choice probability of the form $g(x'\beta)$, which is general enough to include all static discrete choice models we consider in Section 2 and we extend to a class of functions that allow for dynamic discrete choice models in Section 5. An important implication of the linear index is that the term inside the model choice probability is monotonic in each element of x that has random coefficients. This monotonicity is exploited in the proof of the theorem. In the theorem the conditions (i) and (ii) are typically assumed in the models we consider, so we need to verify only the condition of g being real analytic. Real analytic functions include (e.g.) polynomials, exponential functions, and logit-type functions. A formal definition of real analytic function is given as

Definition 1. A function $g(t)$ is **real analytic** at $c \in \mathcal{T} \subseteq \mathbb{R}$ whenever it can be represented as a convergent power series, $g(t) = \sum_{d=0}^{\infty} a_d (t-c)^d$, for a domain of convergence around c. The function $g(t)$ is **real analytic** on an open set $\mathcal{T} \subseteq \mathbb{R}$ if it is real analytic at all arguments $t \in \mathcal{T}$.

Theorem 1 implies that $\lim_{n\to\infty} \rho(F_n, F_0) = 0$ if and only if $\lim_{n\to\infty} d(G_n, G_0) = 0$. Note that $\lim_{n\to\infty} \rho(F_n, F_0) = 0$ implies $\lim_{n\to\infty} d(G_n, G_0) = 0$ is obvious when the convergence in the metric ρ is equivalent to the weak convergence of measures. For example, this holds for the Lévy-Prokhorov metric if the metric space $(\mathcal{B}, \tau)$ is separable where τ is a metric on the set $\mathcal{B}$. Theorem 1 implies that the opposite is also true as long as $\mu(\mathcal{X}) \neq 0$. Therefore this identification result is also useful to show the consistency of a sieve approximation based estimator of F_0 as in Fox, Kim, and Yang (2013). Also note that in Theorem 1 we do not require $\mathcal{X} = \mathbb{R}^K$. Therefore our identification result is different from the identification at infinity and is also different from the Cramer-Wold device.

3.2 Identification with fixed coefficients

Note that when a subset (at least one) of coefficients are not random, then the identification of the distribution of random coefficients is also obtained because we can let

$$h(x, \beta) = g(x_1'\beta_1 + x_2'\beta_2)$$

and treat this is affine in β_2 where β_1 is fixed parameters and β_2 are random coefficients. Our identification strategy of this case applies in two stages. The identification of homogenous coefficients is trivial when x_2 can take the value of zero. At $x_2 = 0$, the model becomes discrete choice models with homogeneous parameters only and their identification is a standard problem. To give further details note that in a first stage of an auxiliary argument we identify the true β_1^0 using the relationship from (9) as

$$G_0(x_1, 0) = \int h(x_1, 0, \beta_1, \beta_2)\, dF_0(\beta_2) = \int g(x_1'\beta_1) dF_0(\beta_2) = g(x_1'\beta_1).$$

Then because $G_0(x_1, 0)$ is known we can identify β_1^0 from the inverse function of the relationship above, typically using a regression. Therefore, in this case we can treat β_1^0 as being known and focus on the identification of $F_0(\beta_2)$. Then the identification of $F_0(\beta_2)$ follows from the corollary below:

Corollary 1. Let $\mathcal{H}_{g_A} = \{h : h(x, \beta) = g(x_1'\beta_1 + x_2'\beta_2), x \in \mathcal{X}\}$ where (i)$\mathcal{X} \subset \mathbb{R}^K$ is a nonempty open set, (ii) includes values of the form$\{(x_1, 0)\}$, and (iii) g is real analytic. Suppose β_1 are fixed coefficients and the support of $F(\beta_2)$, $\mathcal{B}_2$ is known. Then $\mathcal{H}_{g_A}$ is generically totally distinguishing if and only if g is non-polynomial. Moreover, $\mathcal{H}_{g_A}$ is also totally distinguishing.

Proof. Corollary 1 is a direct application of Theorem 1 or Lemma 3.7 in Stinchcombe and White (1998) because $g(x_1'\beta_1 + x_2'\beta_2)$ is affine in β_2 given β_1. □

Below we focus on the models with random coefficients only because all theorems we develop will apply to the models with a subset of fixed parameters after a first stage of identifying the fixed parameters is applied.

3.3 Identification with unknown support of the distribution

Often we do not know the support of F, $\mathcal{B}$. For this reason, it will be useful to strengthen the identification result when the mixture in (9) is generated by any compact subset $\mathcal{B}$. We define this stronger notion of identification as

Definition 2. $\mathcal{H}$ is completely distinguishing if it is totally distinguishing for any distribution $F(\neq F_0) \in \mathcal{F}$ supported on any compact $\mathcal{B}$.

The implication of $\mathcal{H}$ being completely distinguishing is that F_0 is identified on any compact support $\mathcal{B}$ while the support of x $\mathcal{X}$ is particularly given. We apply this notion of identification to the class of functions $\mathcal{H}_g = \{h : h(x, \beta) = g(x'\beta), x \in \mathcal{X}\}$.

As discussed in Stinchcombe and White (1998) whether $\mathcal{H}_g$ is totally distinguishing is equivalent to whether the linear span of $\mathcal{H}_g$ defined below is weakly dense in $C(\mathcal{B})$. We define the linear spaces of functions, spanned by $\mathcal{H}_g$ as

$$\Sigma(\mathcal{H}_g, \mathcal{X}, \mathcal{B}) = \left\{ \begin{array}{c} h : \mathcal{B} \to \mathbb{R} | h(\beta) = \gamma_0 + \sum_{l=1}^{L} \gamma_l g(x^{(l)\prime}\beta), \gamma_0, \gamma_l \in \mathbb{R}, \\ x^{(l)} \in \mathcal{X} \subset \mathbb{R}^K, l = 1, \ldots, L. \end{array} \right\}$$

When $\mathcal{X} = \mathbb{R}^K$, the totally distinguishing property is not surprising. More interesting result is obtained when $\mathcal{X}$ is a subset of $\mathbb{R}^K$. In the proof of Theorem 3 below, we show that $\Sigma(\mathcal{H}_g, \mathcal{X}, \mathcal{B})$ is weakly dense in $C(\mathcal{B})$ and so the identification result follows also with any nonempty open subset $\mathcal{X}$ of $\mathbb{R}^K$.

The difference between $\text{sp}\mathcal{H}_g(\mathcal{X})$ and $\Sigma(\mathcal{H}_g, \mathcal{X}, \mathcal{B})$ is that $\text{sp}\mathcal{H}_g(\mathcal{X})$ does not include the constant functions while $\Sigma(\mathcal{H}_g, \mathcal{X}, \mathcal{B})$ does. But the difference disappears when $\mathcal{X}$ includes $\{0\}$ or an x^* such that $g(x^{*\prime}\beta)$ does not depend on β and $g(x^{*\prime}\beta) \neq 0$. Therefore, in this case $\text{sp}\mathcal{H}_g(\mathcal{X})$ becomes dense in $C(\mathcal{B})$, which is our key argument for identification. Stinchcombe and White (1998) achieves the same goal of showing $\text{sp}\mathcal{H}_g(\mathcal{X})$ is equivalent to $\Sigma(\mathcal{H}_g, \mathcal{X}, \mathcal{B})$ by assuming $g(x'\beta)$ is affine in β.

For $\mathcal{H}_g$, it then becomes completely distinguishing when $\Sigma(\mathcal{H}_g, \mathbb{R}^K, \mathcal{B})$ (so $\mathcal{X} = \mathbb{R}^K$) is uniformly dense in $C(\mathcal{B})$ for any compact $\mathcal{B}$. This uniform denseness is satisfied as long as for the non-polynomial function $g(t)$, there exists an interval $t \in [a, b]$ such that g is Riemann integrable in $[a, b]$ and is continuous on $[a, b]$ due to Hornik (1991). Also see Lemma 3.5 in Stinchcombe and White (1998).

Theorem 2. Let $\mathcal{X} = \mathbb{R}^K$ and $\mathcal{H}_g = \{h : h(x, \beta) = g(x'\beta), x \in \mathcal{X}\}$ where g is Riemann integrable and continuous on $\exists[a, b]$ and non-polynomial. Then $\mathcal{H}_g$ is completely distinguishing.

Proof. The theorem follows from $\text{sp}\mathcal{H}_g = \Sigma(\mathcal{H}_g, \mathbb{R}^K, \mathcal{B})$ and by Lemma 3.5 in Stinchcombe and White (1998). □

Theorem 2 show that a wide class of functions g - that include all of the discrete choice models we consider in Section 2 - can identify the distribution of random coefficients as long as we have the full support condition, $\mathcal{X} = \mathbb{R}^K$ but this full support condition is very strong requirement. Also note that we have not seen any role of analytic function in the identification because any non-polynomial real analytic function satisfies the requirement on g in Theorem 2. The following theorem shows that we can relax the full support condition for the identification when g is real analytic. This includes exponential functions and more importantly logit functions.

This also reveals the role of analytic function in the identification. It effectively removes the full support requirement, which is very important for discrete choice models where the values of covariates are bounded below and above.

Now we show the above completeness result is generically true for any nonempty open subset $\mathcal{X} \subset \mathbb{R}^K$ when the function g is analytic. We further define

Definition 3. $\mathcal{H}$ is generically completely distinguishing if and only if it is totally distinguishing for any open set $\mathcal{X}$ with nonempty interior and for any distribution $F(\neq F_0) \in \mathcal{F}$ supported on any compact $\mathcal{B}$.

Theorem 3. $\mathcal{H}_g$ is generically completely distinguishing when $\{0\} \subset \mathcal{X}$ if and only if g is real analytic and is not a polynomial.

Proof. See Section 3.4 for the proof. □

Theorem 3 is most general and is our main theorem. Note that this identification result also holds for models with a subset of coefficients (at least one) being not random because all theorems we develop apply to the models with a subset of fixed parameters after a first stage of identifying the fixed parameters is applied.

Corollary 2. Let $\mathcal{H}_{g_A} = \{h : h(x, \beta) = g(x_1'\beta_1 + x_2'\beta_2), x \in \mathcal{X}\}$ where g is a real analytic non-polynomial function, $\{(x_1, 0)\} \subset \mathcal{X}$, and β_1 are fixed coefficients. Then $\mathcal{H}_{g_A}$ is generically completely distinguishing.

Proof. See Appendix B for the proof. □

3.4 Proof of Theorem 3

Because Theorem 3 implies Theorem 1 with the known support $\mathcal{B}$, we only prove Theorem 3. We first show that for $\mathcal{H}_g$, the generic completeness is equivalent to the condition that for every $\mathcal{X}$ with nonempty interior, $\Sigma(\mathcal{H}_g, \mathcal{X}, \mathcal{B})$ is uniformly dense in $C(\mathcal{B})$ for any compact $\mathcal{B}$.

Lemma 1. The class $\mathcal{H}_g$ is generically completely distinguishing if and only if for every open set $\mathcal{X}$ with nonempty interior, $\Sigma(\mathcal{H}_g, \mathcal{X}, \mathcal{B})$ is uniformly dense in $C(\mathcal{B})$ for any compact $\mathcal{B}$.

In the proof we use the fact that $x'\beta$ is monotonic in each element of x. By construction of the linear index, this monotonicity is trivially satisfied for the static discrete choice models. For the dynamic discrete choices, to apply the theorem, we need to verify the choice specific continuation payoffs function $EV(x, d; \beta, \alpha)$ is monotonic in each element of x and we verify this in Section 5.

Then, we show that

Lemma 2. $\mathcal{H}_g$ is generically completely distinguishing if and only if it is completely distinguishing when g is real analytic.

The most important implication of Lemma 2 is that we have only to show the identification at a particular choice of $\mathcal{X}$. According to Lemma 2, then the identification must also hold for any $\mathcal{X}$ with nonempty interior. This result facilitates applications of the identification argument substantially because one can take the full support $\mathcal{X} = \mathbb{R}^K$ under which the identification is often easier to show (see e.g. Fox, Kim, Ryan, and Bajari 2012). Note that verifying identification with $\mathcal{X} = \mathbb{R}^K$ does not mean we indeed require the true data should have the full support. It only means that if one shows identification as if $\mathcal{X} = \mathbb{R}^K$, then identification must also hold for any $\mathcal{X}$ with nonempty interior, which includes the real data situation.

Finally combining Lemma 1 and 2 we conclude that Theorem 3 holds because $\mathcal{H}_g$ is completely distinguishing as long as g is real analytic by Theorem 2. In the appendix we prove Lemma 1 and Lemma 2.

3.5 Identification with non-analytic functions

We also find that the class of functions that is generically completely distinguishing is not limited to analytic functions. Other class of functions that satisfy the following condition is also generically completely distinguishing. This includes the normal cumulative distribution function. Therefore the distribution of random coefficients in the probit model is also nonparametrically identified.

Theorem 4. Suppose that $\text{sp}\{d^p g(t), 0 \leq p < \infty | t \in \mathcal{T}\}$ is dense in $C(\mathbb{R})$ for any nonempty open subset $\mathcal{T} \subset \mathbb{R}$ containing $\{0\}$ with $g(\cdot)$ infinitely differentiable. Then for any open set $\mathcal{X} \subset \mathbb{R}^K$ with nonempty interior, the span $\Sigma(\mathcal{H}_g, \mathcal{X}, \mathcal{B})$ is uniformly dense in $C(\mathcal{B})$ for any compact $\mathcal{B}$, so $\mathcal{H}_g$ is generically completely distinguishing.

Proof. Theorem 4 trivially follows from Theorem 3.10 in Stinchcombe and White (1998). □

4 Identification for Static Discrete Choice Models

We verify identification conditions for the examples of static discrete choice models we consider in Section 2. Because other conditions for identification are either trivially satisfied or can be directly assumed, we focus on showing the type specific model choice probability function is either being real analytic - as the key condition in Theorem 3 - or belongs to other class of generically completely distinguishing as in Theorem 4.

4.1 Logit model with individual choices

For the multinomial logit model (2) our identification argument on $F(\beta)$ proceeds after we recover the constant term α from a first stage using an auxiliary argument that does not depend on β. A similar strategy was used in Fox, Kim, Ryan, and Bajari (2012) to identify homogenous parameters in a first stage. The typical strategy is using the observed choice probability at $x_i = 0$ where we have $P(y_{i,j} = 1 | x_i = 0) = \frac{\exp(\alpha)}{1+J\exp(\alpha)}$. Because $P(y_{i,j} = 1 | x_i = 0)$ is nonparametrically identified from the data, α is also identified from the inverse function. Below we focus on the identification of $F(\beta)$ assuming α is identified in the first stage. With abuse of notation we write $g_j(x_i, \beta) = g_j(x_i, \beta, \alpha^0)$ where α^0 denotes the true value of α.

In Section 3 we have shown that the key identification condition of $F(\beta)$ is that (i) $g_j(x_i,\beta)$ is real analytic, (ii) the support of distribution x_i, $\mathcal{X}$ has nonempty interior, and (iii) $\mathcal{X}$ includes at least one value of $x=\tilde{x}$ such that $g_j(\tilde{x},\beta)\neq 0$ does not depend on β. We assume the condition (ii). To satisfy the condition (iii) simply we can take $\tilde{x}=0$ in static discrete choice models where the covariates are re-centered at zero. For a binary logit model, it is obvious that $g_j(x_i,\beta)$ is real analytic, so the condition (i) is also satisfied. For the multinomial logit case too we can show that $g_j(x,\beta)$ is real analytic as follows. Pick a particular j and let $x_{j'}=0$ for all $j'\neq j$. Then we have $g_j(x_i,\beta)=\frac{\exp(x'_{i,j}\beta)}{\exp(-\alpha^0)+J-1+\exp(x'_{i,j}\beta)}$, which has the form $G_\eta(t)\equiv\frac{\exp(t)}{\eta+\exp(t)}$ for some η and hence $g_j(x_i,\beta)$ is real analytic because the exponential function is real analytic and the function $g_j(x_i,\beta)$ is formed by the addition and division of never zero real analytic functions (Krantz and Parks 2002).

4.2 Nested logit model with individual choices

First we show ρ_j - that reflects the correlation between goods for each group - is identified from an auxiliary step. Note that where $z_i=0$ and $x_i=0$, we have

$$P^0_{j,l}\equiv P(y_{i,j,l}=1|z_i=0,x_i=0)=g_{j,l}(0,0,\gamma,\beta,\rho)=\frac{\exp\left(\rho_j\log\left(L_j\right)\right)}{\sum_{j'=0}^{J}\exp\left(\rho_{j'}\log\left(L_{j'}\right)\right)}\frac{1}{L_j}.$$

It follows that $L_j\cdot P^0_{j,l}=\frac{\exp(\rho_j\log(L_j))}{\sum_{j'=0}^{J}\exp\left(\rho_{j'}\log\left(L_{j'}\right)\right)}$ and therefore $\log(L_j\cdot P^0_{j,l})-\log(L_0\cdot P^0_{0,l})=\rho_j\log(L_j)$, from which we identify ρ_j for $j=1,\ldots,J$ as

$$\rho_j=\{\log(L_j\cdot P^0_{j,l})-\log(L_0\cdot P^0_{0,l})\}/\log(L_j)$$

because $P^0_{j,l}$ and L_j are directly observable from data for all j,l. Below we treat ρ_j's as known.

In the nested logit model of (3) we focus on showing $g_{j,l}(z_i,x_i,\gamma,\beta,\rho)$ is a real analytic function. Other conditions for identification in Section 3 are trivially satisfied or directly assumed as in the multinomial logit case.

Now pick a particular j and let $z_{j'}=0$ for all $j'\neq j\in\{0,1,\ldots,J\}$ and let $x_{j,l}=0$ for all j and l. Then we have

$$\begin{aligned}g_{j,l}(z_i,x_i,\gamma,\beta,\rho)&=\frac{\exp\left(z'_{i,j}\gamma+\rho_j\log(L_j)\right)}{\sum_{j'\neq j}\exp\left(\rho_{j\prime}\log\left(L_{j\prime}\right)\right)+\exp\left(z'_{i,j}\gamma+\rho_j\log(L_j)\right)}\frac{1}{L_j}\\&=\frac{\exp\left(z'_{i,j}\gamma\right)}{\sum_{j'\neq j}\exp\left(\rho_{j\prime}\log\left(L_{j\prime}\right)-\rho_j\log\left(L_j\right)\right)+\exp\left(z'_{i,j}\gamma\right)}\frac{1}{L_j}=\frac{\exp\left(z'_{i,j}\gamma\right)}{\eta+\exp\left(z'_{i,j}\gamma\right)}\frac{1}{L_j}\end{aligned}$$

where we let $\eta=\sum_{j'\neq j}\exp\left(\rho_{j\prime}\log\left(L_{j\prime}\right)-\rho_j\log\left(L_j\right)\right)$.

Therefore $g_{j,l}(z_i,x_i,\gamma,\beta,\rho)$ has the form as $\frac{1}{L_j}G_\eta(t)=\frac{\exp(t)}{\eta+\exp(t)}\frac{1}{L_j}$, so is an analytic function as in the multinomial logit case (Krantz and Parks 2002). Therefore the distribution of the random coefficients γ is identified when z_j also includes $\{0\}$. Now we turn to the identification of the distribution of β_j. Let $z_j = 0$ for all j and let $x_{j',l'} = 0$ for all $j' \neq j$ and $l' \neq l$. Then we have

$$
\begin{aligned}
& g_{j,l}(z_i, x_i, \gamma, \beta, \rho) \\
=& \frac{\exp\left(\rho_j \log\left(L_j - 1 + \exp(x'_{i,j,l}\beta_j/\rho_j)\right)\right)}{\sum_{j' \neq j} \exp\left(\rho_{j'} \log\left(L_{j'}\right)\right) + \exp\left(\rho_j \log\left(L_j - 1 + \exp(x'_{i,j,l}\beta_j/\rho_j)\right)\right)} \\
& \times \frac{\exp(x'_{i,j,l}\beta_j/\rho_j)}{L_j - 1 + \exp(x'_{i,j,l}\beta_j/\rho_j)} \\
=& \frac{\left(L_j - 1 + \exp(x'_{i,j,l}\beta_j/\rho_j)\right)^{\rho_j}}{\sum_{j' \neq j} \exp\left(\rho_{j'} \log\left(L_{j'}\right)\right) + \left(L_j - 1 + \exp(x'_{i,j,l}\beta_j/\rho_j)\right)^{\rho_j}} \frac{\exp(x'_{i,j,l}\beta_j/\rho_j)}{L_j - 1 + \exp(x'_{i,j,l}\beta_j/\rho_j)}
\end{aligned}
$$

Because the product of analytic functions is also analytic, we have only to show the function

$$
\widetilde{G}_{\widetilde{\eta}}(t) = \frac{(L_j - 1 + \exp(t/\rho_j))^{\rho_j}}{\widetilde{\eta} + (L_j - 1 + \exp(t/\rho_j))^{\rho_j}}
$$

(where we write $\widetilde{\eta} = \sum_{j' \neq j} \exp\left(\rho_{j'} \log\left(L_{j'}\right)\right)$) is analytic because $\frac{\exp(x'_{i,j,l}\beta_j/\rho_j)}{L_j - 1 + \exp(x'_{i,j,l}\beta_j/\rho_j)}$ is analytic (it can be written as $\frac{\exp(t)}{\eta + \exp(t)}$ for some η). $\widetilde{G}_{\widetilde{\eta}}(t)$ is also analytic as long as $(L_j - 1 + \exp(t/\rho_j))^{\rho_j}$ is analytic because the reciprocal of an analytic function that does not take the value of zero at its support is also analytic. Now note that $(L_j - 1 + \exp(t/\rho_j))^{\rho_j}$ is analytic because compositions of analytic functions are analytic and $L_j - 1 + \exp(t/\rho_j)$ is strictly positive. Therefore we conclude the distribution of random coefficients of β_j is identified. Similarly we can show that all the distributions of β_j, $j = 1, \ldots, J$ are identified.

4.3 Probit model with binary choice

We assume the support of distribution of $x_{i,1}$ includes $\{0\}$ (or re-centered at zero). In a first stage we identify α^0 from $P(y_{i,1} = 1 | x_{i,1} = 0) = \Phi(\alpha^0)$ at $x_{i,1} = 0$. Also $\Phi(\cdot)$ does not depend on β at $x_{i,1} = 0$ and $\Phi(\alpha^0) \neq 0$. Finally although the normal CDF $\Phi(\cdot)$ is not analytic, it is infinitely differentiable and satisfies conditions in Theorem 4. Therefore the distribution $F(\beta)$ is identified in this case too.

4.4 Logit model with aggregate data

As the logit model with individual choices, by the similar argument, the distribution of random coefficients is identified. The only difference is that in the individual choices we identify $P(y_{i,j} = 1 | x_i)$ from the data in an auxiliary step while in the aggregate data case, the conditional share s_j is the data.

5 Identification for Dynamic Programming Discrete Choices

We have shown that the distribution of random coefficients is identified for static discrete choice models. However, the theorems in Section 3 cannot be directly applied to the dynamic programming discrete choice problems because the type specific model choice probabilities in these models contain the choice specific continuation payoffs functions. In this section first we show that the choice specific continuation payoffs function and so the choice specific value function is monotonic in each element of covariates vector that have random coefficients. Then we show that Theorem 3 can extend to the dynamic programming discrete choice problems based on this monotonicity result.[2]

Following Rust (1994), let $u(x, d, \beta, \alpha)$ denote the per period utility of taking an action d in the set of choices $D(x)$ where x denotes the covariates or states variables with random coefficients, β denotes random coefficients, and α denotes homogeneous coefficients. Let $EV(x, d; \beta, \alpha)$ denote the choice specific continuation payoffs function or the choice specific expected value function. Then for the logit model the type specific choice probability of taking the action d becomes

$$g_d(x, \beta, \alpha) = \frac{\exp\{u(x, d, \beta, \alpha) + \delta EV(x, d; \beta, \alpha)\}}{\sum_{d' \in D(x)} \exp\{u(x, d', \beta, \alpha) + \delta EV(x, d'; \beta, \alpha)\}}$$

where the expected value function $EV(x, d; \beta, \alpha)$ of the logit model is given by the unique fixed point that solves

$$EV(x, d; \beta, \alpha) = \int_y \log \left\{ \sum_{d' \in D(y)} \exp\{u(y, d', \beta, \alpha) + \delta EV(y, d'; \beta, \alpha)\} \right\} \pi(dy|x, d)$$

where $\pi(dy|x, d)$ denotes the transition density depending on d. We note that

Lemma 3. Suppose the per period utility satisfies the linear index restriction, i.e., $u(x, d, \beta, \alpha)$ depends on $x_d'\beta$ but does not depend on x_d or β, separately. Then the choice specific continuation payoffs function $EV(x, d; \beta, \alpha)$ is monotonic in each element of x.

Proof. See Appendix C for the proof. Note that the result and its proof are not specific to the logit model. □

Based on this monotonicity, next we obtain the identification of the distribution of random coefficients for dynamic discrete choice models.

Theorem 5. Let $g_d(x, \beta, \alpha)$ be the type specific choice probability of a dynamic discrete choice problem. Then$\mathcal{H}_g^D = \{h : h = g_d(x, \beta, \alpha), x \in \mathcal{X}\}$ is generically completely distinguishing when $\{0\} \subset \mathcal{X}$ if and only if (i) g_d is real analytic and is not a polynomial and (ii) the per period utility $u(x, d, \beta, \alpha)$ satisfies the linear index restriction in Lemma 3.

[2]Theorem 1 also extends to the dynamic discrete choices because Theorem 3 implies Theorem 1.

Proof. See Appendix D for the proof. □

In the example of the dynamic binary choice model of (5)-(6), $\mathcal{H}_g^D$ becomes

$$\mathcal{H}_g^D = \left\{ h : h = g_1(x, \beta, \alpha) = \frac{\exp\{x'\beta + \delta EV(x, 1; \beta, \alpha)\}}{\exp\{\alpha + \delta EV(x, 0; \beta, \alpha)\} + \exp\{x'\beta + \delta EV(x, 1; \beta, \alpha)\}}, x \in \mathcal{X} \right\}$$

and we prove the identification theorem for the binary case in the appendix without loss of generality because for the multinomial choices, we can let $x_d = 0$ for $d \neq 1$. In the proof we assume the discount factor δ, the scrap value α, and the transition density $\pi(dy|x, d)$ are known due to the following remark:

Remark 1. Rust (1987, 1994) and Magnac and Thesmar (2002) argue that it is difficult to identify the discount factor δ, so we assume it is known. For the binary logit case the homogeneous parameter, scrap value α is identified at $x = 0$ from the observation that $P(1|x = 0) = \int g_1(0, \beta, \alpha) dF(\beta) = \int \frac{1}{\exp\{\alpha\}+1} dF(\beta) = \frac{1}{\exp\{\alpha\}+1}$ because from (7) we find $EV(0, 1; \beta, \alpha) = EV(0, 0; \beta, \alpha)$ (see also Rust 1987) - which is obvious because when $x = 0$ it does not matter whether the bus engine is new or not. The transition density $\pi(dy|x, d)$ is also nonparametrically identified from the data. Therefore, we can focus on the identification of the distribution of random coefficients.

6 Conclusion

We show that the distributions of random coefficients in various discrete choice models are nonparametrically identified. Our identification results apply to both binary and multinomial logit, nested logit, and probit models as well as dynamic programming discrete choices. To our best knowledge, this is the first formal result to show the nonparametric identification of random coefficients in the dynamic discrete choices.

We find that the distribution of random coefficients is identified if (i) the type specific model choice probability belongs to a class of functions that include real analytic functions and the support of the distribution of covariates is a nonempty open set, (ii) the term inside the type specific choice probability is monotonic in each element of the covariates vector that has random coefficients, and (iii) the type specific choice probability does not depend on random coefficients at a particular value of covariates. We show that these conditions are satisfied for various discrete choice models that are commonly used in the empirical studies. In our identification we stress the role of analytic function that effectively removes the full support requirement often exploited in other identification approaches, which is very important for discrete choice models where the values of covariates are often bounded below and above.

Lastly, as a referee points out, our identification results can be used as basis for specification testing. First note that our identification allows for the case of degenerated distribution (i.e., coefficients are fixed parameters, not random) and hence can serve as a specification test for a random coefficient model. Moreover, our results can be used as a specification test for the specific choice model. Suppose a known $\psi(x,\beta)$, with appropriate normalization, is incorrectly used instead of the true $h(x,\beta)$ in (9). Then since

$$
\begin{aligned}
G_0(x) &= \int h(x,\beta)dF_0(\beta) = \int \psi(x,\beta)\frac{h(x,\beta)}{\psi(x,\beta)}dF_0(\beta) \\
&= \int \psi(x,\beta)\, dH(x,\beta)
\end{aligned}
$$

for an H such that $dH(x,\beta) = \frac{h(x,\beta)}{\psi(x,\beta)}dF_0(\beta)$, if $\psi(x,\beta)$ satisfies the identification conditions, then any distribution function of β only - which is not a function of x - will be rejected from our identification exercise. Therefore one can conclude$\psi(x,\beta)$ is incorrectly specified. Although these specification tests seem promising based on our identification results, a formal development should be addressed with further research.

Appendix

A Proof of Lemma's for Theorem 3

A.1 Proof of Lemma 1

Our proof will closely follow the proof of Lemma 3.7 in Stinchcombe and White (1998). We prove Lemma 1 for dynamic discrete choice models in Section D. Lemma 1 for the static discrete choice models with $\mathcal{H}_g$ can be proved by the essentially same arguments by taking the discount factor $\delta = 0$ i.e. we drop the continuation payoffs function $EV(x, d; \beta, \alpha)$ in the type specific model choice probability.

A.2 Proof of Lemma 2

If $\mathcal{H}_g$ is generically completely distinguishing, it is also completely distinguishing by definitions. Next we show the opposite is also true. If $\mathcal{H}_g$ is not generically completely distinguishing, we can find a compact set $\widetilde{\mathcal{B}}$ and a nonempty open set $\widetilde{\mathcal{X}}$ such that $\Sigma(\mathcal{H}_g, \widetilde{\mathcal{X}}, \widetilde{\mathcal{B}})$ is not uniformly dense in $C(\widetilde{\mathcal{B}})$. Then there exists a distribution $\widetilde{F} \neq F_0$ supported on $\widetilde{\mathcal{B}}$ such that for all $x \in \widetilde{\mathcal{X}}$, $\widetilde{G}(x) = \int g(x'\beta)\, d\left(\widetilde{F}(\beta) - F_0(\beta)\right) = 0$ by the Hahn-Banach theorem. We, however, note that $\widetilde{G}(x)$ is real analytic because $g(\cdot)$ is and $\widetilde{\mathcal{B}}$ is compact. We further note that a real analytic function is equal to zero on the open set $\widetilde{\mathcal{X}}$ if and only if it is equal to zero everywhere. This implies that $\mathcal{H}_g$ is not completely distinguishing. Therefore if $\mathcal{H}_g$ is completely distinguishing, it must be also generically completely distinguishing. This completes the proof.

In the proof $\widetilde{G}(x)$ is a multivariate function. According to Definition 2.2.1 in Krantz and Parks (2002) a function $\Delta(x)$, with domain an open subset $\mathcal{T} \subseteq \mathbb{R}^K$ and range $\mathbb{R}$, is called (multivariate) real analytic on $\mathcal{T}$ if for each $x \in \mathcal{T}$ the function $\Delta(\cdot)$ may be represented by a convergent power series in some neighborhood of x.

B Proof of Corollary 2

This can be proved similarly with the proof of Theorem 3 or the proof of Lemma 3.7 in Stinchcombe and White (1998).

C Proof of Lemma 3

Proof. We prove the lemma for the dynamic binary choice without loss of generality. Let $\check{x}^{(k)}$ be a vector of states that is equal to x except the k-th element. Let $E\tilde{V}(\check{x}, d; \beta, \alpha)$ denote the value function when an agent having the covariates or states equal to $\check{x}^{(k)}$ takes a sequence of choices that are optimal under the current state x. Without loss of generality we consider the case that β_k, the k-th element in β is positive and $\check{x}^{(k)} \geq x$. Then we have

$$E\tilde{V}(\check{x}^{(k)}, d; \beta, \alpha) \leq EV(\check{x}^{(k)}, d; \beta, \alpha)$$

because $EV(\check{x}^{(k)}, d; \beta, \alpha)$ is the value of the expected value function when an agent with the states equal to $\check{x}^{(k)}$ takes a sequence of optimal choices by the definition of the value function and $E\tilde{V}(\check{x}^{(k)}, d; \beta, \alpha)$ is from a non-optimal choices of actions. Next we note that

$$EV(x, d; \beta, \alpha) \leq E\tilde{V}(\check{x}^{(k)}, d; \beta, \alpha)$$

because (i) for any time period the per period utility under $\check{x}^{(k)}$ is greater than or equal to the per period utility under x and (ii) the agent takes the same sequence of choices under x and $\check{x}^{(k)}$ in our definition of $E\tilde{V}(\check{x}^{(k)}, d; \beta, \alpha)$. Combining these two results, we conclude the monotonicity because

$$EV(x, d; \beta, \alpha) \leq EV(\check{x}^{(k)}, d; \beta, \alpha) \text{ whenever } \check{x}_k^{(k)} \geq x_k.$$

Our choice of the k-th element is arbitrary and so this monotonicity result holds for any element in x. □

D Proof of Theorem 5

We prove this theorem by showing corresponding results to Lemma 1 and Lemma 2 hold for $\mathcal{H}_g^D$. Lemma 2 holds trivially since the function g in $\mathcal{H}_g^D$ is analytic. We focus on Lemma 1. We prove this for the dynamic programming binary choice model of (5)-(6) without loss of generality. We assume the discount factor δ is known. We also assume α is known since it can be identified from an auxiliary step as discussed in Remark 1.

Define the linear spaces of functions, spanned by $\mathcal{H}_g^D$ as

$$\Sigma(\mathcal{H}_g^D, \mathcal{X}, \mathcal{B}) = \left\{ \begin{array}{c} h : \mathcal{B} \to \mathbb{R} | h(\beta) = \gamma_0 + \sum_{l=1}^{L} \gamma_l g(x^{(l)}, \beta, \alpha), \gamma_0, \gamma_l \in \mathbb{R}, \\ x^{(l)} \in \mathcal{X} \subset \mathbb{R}^K, l = 1, \ldots, L. \end{array} \right\}.$$

If the uniform closure of $\text{sp}\mathcal{H}_g^D(\mathcal{X})$ contains $C(\mathcal{B})$, then that of $\Sigma(\mathcal{H}_g^D, \mathcal{X}, \mathcal{B})$ also must contain $C(\mathcal{B})$ since $\text{sp}\mathcal{H}_g^D(\mathcal{X}) \subset \Sigma(\mathcal{H}_g^D, \mathcal{X}, \mathcal{B})$ by construction. Now suppose the uniform closure of $\Sigma(\mathcal{H}_g^D, \mathcal{X}, \mathcal{B})$ contains $C(\mathcal{B})$ for every compact $\mathcal{B} \subset \mathbb{R}^K$ and suppose that $\mathcal{X} \subset \mathbb{R}^K$ has nonempty interior containing $\{0\}$. We will prove Theorem 5 by contradiction. We prove this for the dynamic programming binary choice problem (say $D = \{0, 1\}$) without loss of generality because for the multinomial choices, we can let $x_d = 0$ for $d \neq 1$. We take $g = g_1$ and let $x = x_1$ below.

Now suppose that $\text{sp}\mathcal{H}_g^D(\mathcal{X})$ is not dense in $C(\mathcal{B})$ for some $\mathcal{X}$ and $\mathcal{B}$. This happens if and only if there exists a distribution function $F \neq F_0$ (in the sense that $\rho(F, F_0) \neq 0$) supported on $\mathcal{B}$ such that for all $x \in \mathcal{X}$, $\int g(x, \beta, \alpha) d\left(F(\beta) - F_0(\beta)\right) = 0$.

Let A be a compact subset of $\mathbb{R}^K$ containing an ϵ-neighborhood of $\mathcal{B}$ (in terms of the Hausdorff metric) for some $\epsilon > 0$. Pick $\delta > 0$ and $\tilde{x} \in \mathcal{X}$ such that $S(\tilde{x}, 2\delta)$, the ball of radius 2δ around $\tilde{x}$, is contained in $\mathcal{X}$. By assumption, $\Sigma(\mathcal{H}_g^D, S(\tilde{x}, 2\delta), A)$ is uniformly dense in $C(A)$. It follows that for every $n \in \mathbb{N}$ and for every strict subset $\tilde{A} \subset A$, some element of $\Sigma(\mathcal{H}_g^D, S(\tilde{x}, \delta), A)$ is uniformly within n^{-1} of the continuous function

$$f^n(\beta) := \max\{1 - n\varrho(\beta, \tilde{A}), 0\}$$

where $\varrho(\beta, \tilde{A})$ is the Hausdorff distance from β to the set $\tilde{A}$. By construction the sequence $f^n(\beta)$ is uniformly bounded between zero and one and converges pointwise to the indicator function $1\left\{\beta \in \tilde{A}\right\}$. Therefore, as n goes to infinity, $\int_A f^n(\beta) d\left(F(\beta) - F_0(\beta)\right)$ goes to $\int_{\tilde{A}} 1 d\left(F(\beta) - F_0(\beta)\right)$. Because each f^n is in the span of $\mathcal{H}_g^D(S(\tilde{x}, \delta))$ and 1, we can write

$$\begin{aligned} f^n(\beta) &= \gamma_{0,n} + \sum_{l,n=1}^{L,n} \gamma_{l,n} g(x^{(l,n)}, \beta, \alpha) \qquad (11) \\ &= \gamma_{0,n} + \sum_{l,n=1}^{L,n} \gamma_{l,n} \frac{\exp\{x^{(l,n)\prime}\beta + \delta\left[EV(x^{(l,n)}, 1; \beta, \alpha) - EV(x^{(l,n)}, 0; \beta, \alpha)\right]\}}{\exp\{\alpha\} + \exp\{x^{(l,n)\prime}\beta + \delta\left[EV(x^{(l,n)}, 1; \beta, \alpha) - EV(x^{(l,n)}, 0; \beta, \alpha)\right]\}} \end{aligned}$$

where each $x^{(l,n)} \in S(\tilde{x}, \delta)$. The key idea underlying this proof strategy is that we can stretch out the functions f^n without changing their integral against $F(\beta) - F_0(\beta)$, and then we show this cannot happen unless $F(\beta) = F_0(\beta)$ for almost all $\beta \in \mathcal{B}$.

Now we formalize the idea. Because $\int g(x, \beta, \alpha) d\left(F(\beta) - F_0(\beta)\right)$ equal to zero for any element $g \in \mathcal{H}_g^D(\mathcal{X})$, we can let any $\check{x}^{(l,n)}$ substitute each $x^{(l,n)}$ in (11) without changing the integral of $f^n(\beta)$ against $F(\beta) - F_0(\beta)$. Without loss of generality we can take $\tilde{A}$ as a Cartesian product of intervals $\prod_{k=1}^K \left[\underline{\beta}_k, \overline{\beta}_k\right]$. Then, for each of K elements we can find a sequence of $b_k^{l,n}$ and $c_k^{l,n}$, $k = 1, \ldots, K$, in $\mathbb{R}^K$ such that

$$x^{(l,n)\prime} b_k^{(l,n)} + \delta\left[EV(x^{(l,n)}, 1; b_k^{(l,n)}, \alpha) - EV(x^{(l,n)}, 0; b_k^{(l,n)}, \alpha)\right] = \underline{\beta}_k$$

and

$$x^{(l,n)\prime} c_k^{(l,n)} + \delta\left[EV(x^{(l,n)}, 1; c_k^{(l,n)}, \alpha) - EV(x^{(l,n)}, 0; c_k^{(l,n)}, \alpha)\right] = \overline{\beta}_k.$$

Because (i) $S(\tilde{x}, \delta) \subset S(\tilde{x}, 2\delta) \subset \mathcal{X}$ and (ii) the function $x'\beta + \delta\left[EV(x, 1; \beta, \alpha) - EV(x, 0; \beta, \alpha)\right]$ is monotonic in each element of x,[3] now we can find some $\eta_k \in (0, \epsilon)$, $k = 1, \ldots, K$ such that for

[3] Lemma 3 implies that the difference of the expected value function, $EV(x, 1; \beta, \alpha) - EV(x, 0; \beta, \alpha)$ in (6) is monotonic in each element of x because $EV(x, 0; \beta, \alpha)$ does not depend on x (Recall that "$d = 0$" denotes the replacement of a bus engine). It also follows that the function $x'\beta + \delta\left[EV(x, 1; \beta, \alpha) - EV(x, 0; \beta, \alpha)\right]$ in (6) is monotonic in each element of x.

all (l, n)-pairs there exists $\check{x}^{(l,n)} \in \mathcal{X}$ such that

$$\check{x}^{(l,n)\prime} b_k^{(l,n)} + \delta \left[EV(\check{x}^{(l,n)}, 1; b_k^{(l,n)}, \alpha) - EV(\check{x}^{(l,n)}, 0; b_k^{(l,n)}, \alpha) \right] = \underline{\beta}_k - \eta_k$$

and

$$\check{x}^{(l,n)\prime} c_k^{(l,n)} + \delta \left[EV(\check{x}^{(l,n)}, 1; c_k^{(l,n)}, \alpha) - EV(\check{x}^{(l,n)}, 0; c_k^{(l,n)}, \alpha) \right] = \underline{\beta}_k + \eta_k.$$

In the sequence of functions $\{f^n\}$ defined in (11), replace each $x^{(l,n)}$ by the corresponding $\check{x}^{(l,n)}$ and obtain a sequence of functions in $\Sigma(\mathcal{H}_g^D, \mathcal{X}, \mathcal{B})$, say $\{h^n\}$. Then the sequence $\{h^n\}$ converges pointwise to the indicator function $1\left\{\beta \in \tilde{A}_\eta\right\}$ where $\tilde{A}_\eta = \prod_{k=1}^K \left[\underline{\beta}_k - \eta_k, \overline{\beta}_k + \eta_k\right]$. Therefore, we find

$$\int_{\widetilde{A}} 1 d\left(F(\beta) - F_0(\beta)\right) = \int_{\widetilde{A}_\eta} 1 d\left(F(\beta) - F_0(\beta)\right)$$

and this cannot be true unless $F(\beta) = F_0(\beta)$ for almost all $\beta \in \mathcal{B}$ because A contains an ϵ-neighborhood of $\mathcal{B}$. Based on this contradiction, we complete the proof.

References

[1] Berry, S and P. Haile (2010), "Nonparametric Identification of Multinomial Choice Models with Heterogeneous Consumers and Endogeneity", Yale University working paper.

[2] Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Price in Market Equilibrium", *Econometrica*, 63(4), 841-890.

[3] Bierens, H. (1982), "Consistent Model Specification Tests", *Journal of Econometrics,* 26, 323-353.

[4] Bierens, H. (1990), "A Consistent Conditional Moment Test of Functional Form", *Econometrica,* 58, 1443-1458.

[5] Boyd, J and R. Mellman (1980), "Effect of Fuel Economy Standards on the U. S. Automotive Market: An Hedonic Demand Analysis", *Transportation Research B*, 14(5), 367-378.

[6] Briesch, R, P. Chintagunta, and R. Matzkin (2010), "Nonparametric Discrete Choice Models with Unobserved Heterogeneity", *Journal of Business and Economic Statistics,* 28(2), 291-307.

[7] Burda, M., M. Harding, and J. Hausman (2008), "A Bayesian Mixed Logit-Probit for Multinomial Choice", *Journal of Econometrics*, 147(2), 232-246.

[8] Cardell, N. and F. Dunbar (1980), "Measuring the Societal Impacts of Automobile Downsizing", *Transportation Research B*, 14(5), 423-434.

[9] Chiappori, P. and I. Komunjer (2009), "On the Nonparametric Identification of Multiple Choice Models", working paper.

[10] Fox, J. and A. Gandhi (2010), "Nonparametric Identification and Estimation of Random Coefficients in Nonlinear Economic Models", University of Michigan working paper.

[11] Fox, J, K. Kim, S. Ryan, and P. Bajari (2012), "The Random Coefficients Logit Model is Identified", *Journal of Econometrics*, 166(2), 204-212.

[12] Fox, J, K. Kim, and C. Yang (2013), "A Simple Nonparametric Approach to Estimating the Distribution of Random Coefficients in Structural Models", working paper.

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

[14] Hoderlein, S., J. Klemala, and E. Mammen (2010), "Reconsidering the Random Coefficient Model", *Econometric Theory*, 26(3), 804-837.

[15] Hornik, K. (1991), "Approximation capabilities of multilayer feedforward networks", *Neural Networks*, 4, 251-257.

[16] Ichimura, H. and T. Thompson (1998), "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution", *Journal of Econometrics*, 86(2), 269-295.

[17] Krantz, S. and H. Parks (2002), *A Primer on Real Analytic Functions*, Second Edition. Birkhauser.

[18] Lewbel, A. (2000), "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity or Instrumental Variables", *Journal of Econometrics*, 97(1), 145-177.

[19] Magnac, T. and D. Thesmar (2002), "Identifying Dynamic Discrete Decision Processes", *Econometrica*, 70, 801-816.

[20] McFadden, D. (1978), "Modelling the Choice of Residential Location", in *Spatial Interaction Theory and Planning Models*, 75-96, A. Karlquist, L. Lundquist, F. Snickars, and J. W. Weibull et al. (Eds), Amsterdam, New York, North-Holland.

[21] McFadden, D. and K. Train (2000), "Mixed MNL models for discrete response*", Journal of Applied Econometrics*, 15(5), 447-470.

[22] Nevo, A. (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry", *Econometrica*, 69(2), 307-342.

[23] Petrin, A. (2002), "Quantifying the Benefits of New Products: The Case of the Minivan", *Journal of Political Economy*, 110, 705-729.

[24] Rossi, P., G. Allenby, and R. McCulloch (2005), *Bayesian Statistics and Marketing*. West Sussex: John Willy & Sons.

[25] Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, 55(5), 999-1033.

[26] Rust, J. (1994), "Structural Estimation of Markov Decision Processes", in *Handbook of Econometrics*, vol. 4, edited by Robert F. Engle and Daniel L. McFadden. Amsterdam: North-Holland.

[27] Stinchcombe, M. and H. White (1998), "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative", *Econometric Theory*, 14, 295-325.