NBER WORKING PAPER SERIES

THE RANDOM COEFFICIENTS LOGIT MODEL IS IDENTIFIED

Patrick Bajari
Jeremy Fox
Kyoo il Kim
Stephen P. Ryan

Working Paper 14934
http://www.nber.org/papers/w14934

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

Bajari thanks the National Science Foundation, various grants, for generous research support. Fox thanks the National Science Foundation, the Olin Foundation, and the Stigler Center for generous funding. Thanks to helpful comments from Azeem Shaikh. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2009 by Patrick Bajari, Jeremy Fox, Kyoo il Kim, and Stephen P. Ryan. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Random Coefficients Logit Model Is Identified
Patrick Bajari, Jeremy Fox, Kyoo il Kim, and Stephen P. Ryan
NBER Working Paper No. 14934
April 2009
JEL No. C14,C25,L00

ABSTRACT

The random coefficients, multinomial choice logit model has been widely used in empirical choice analysis for the last 30 years. We are the first to prove that the distribution of random coefficients in this model is nonparametrically identified. Our approach exploits the structure of the logit model, and so requires no monotonicity assumptions and requires variation in product characteristics within only an infinitesimally small open set. Our identification argument is constructive and may be applied to other choice models with random coefficients.

Patrick Bajari
Professor of Economics
University of Minnesota
4-101 Hanson Hall
1925 4th Street South
Minneapolis, MN 55455
and NBER
bajari@econ.umn.edu

Jeremy Fox
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
fox@uchicago.edu

Kyoo il Kim
Department of Economics
University of Minnesota
949 Heller Hall
271 19th Ave South
Minneapolis, MN 55455
kyookim@umn.edu

Stephen P. Ryan
MIT Department of Economics
E52-262C
50 Memorial Drive
Cambridge, MA 02142
and NBER
sryan@mit.edu

1 Introduction

In economics, it is common to observe that otherwise identical agents behave differently when faced with identical choice environments, due to such factors as heterogeneity in preferences. A growing econometric literature has addressed this problem by providing estimators that allow the coefficients of the economic model to vary across agents. One of the most commonly used models in applied choice analysis is the random coefficients logit model, which models the decision of a consumer to choose between one of a finite number of competing alternatives. Hausman and Wise (1978) introduced flexible specifications for discrete choice models, while Boyd and Mellman (1980) as well as Cardell and Dunbar (1980) introduced the random coefficients logit model. Since then, the random coefficients logit model has formed the basis for hundreds of empirical studies. The book by Train (2003) calls the random coefficients logit the "mixed logit". An expanded version of the random coefficients logit that deals with aggregate demand shocks was introduced by Berry, Levinsohn and Pakes (1995) and itself has been used in hundreds of studies, including Nevo (2001) and Petrin (2002).

In the random coefficients logit, consumers can choose between $j = 1, ..., J$ mutually exclusive inside goods and one outside good. The exogenous variables for choice j are in the $K \times 1$ vector x_j. In the example of demand estimation, x_j might include the product characteristics of good j and the demographics of the consumer. We shall let $x = (x'_1, ..., x'_J)'$ denote the stacked vector of all the x_j. Each consumer has a preference parameter β, which is a vector of K marginal utilities that give the consumer's preferences over the K product characteristics. The consumer's utility for choice j is equal to

$$u_i = x'_j \beta + \epsilon_j. \tag{1}$$

The outside good has a utility of $u_0 = \epsilon_0$. The logit model is defined when the errors ϵ_j are i.i.d. across choices and each error has the Type I extreme value distribution, which has a CDF of $\exp(-\exp(-\epsilon_j))$.[1] The random coefficients logit arises when β varies across the population, with unknown density $f(\beta)$. The object of identification is the density f. Under the standard assumption that β is independent of x, choice probabilities are

$$\Pr(j \mid x;\, f) = \int \frac{\exp\left(x'_j \beta\right)}{1 + \sum_{h=1}^{J} \exp\left(x'_h \beta\right)} f(\beta)\, d\beta. \tag{2}$$

[1]The Type I extreme value distribution gives the scale normalization for utility values. The outside good's utility gives the location normalization.

This specification is popular with empirical researchers because the resulting choice probabilities are relatively flexible. Let price be in x_j. 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. McFadden and Train (2000) consider a nonparametric choice of f and allow the linear index $x_j'\beta$ to be a flexible polynomial in some underlying product characteristics. They prove that this combination can flexibly approximate any choice probabilities (using the same underlying product characteristics) that arise from a random utility model. McFadden and Train do not study identification.

Previous empirical implementations have made functional form assumptions (often normal) on f, so that f is known up to a finite vector of parameters. We are the first to explore whether the distribution f is nonparametrically identified: whether variation in x is enough so that the true f^0 is the only f that solves (2) for some j and all x. In operator notation, (2) can be written as $P_j = Q_j(f)$. The density f^0 is nonparametrically identified if the choice probability operator Q_j is one-to-one: each density f gives a unique choice-probability (data) function P_j. For the true f^0 and an alternative $f^1 \neq f^0$, Q_j being one-to-one is equivalent to saying there exists an x where $\Pr\left(j \mid x;\, f^0\right) \neq \Pr\left(j \mid x;\, f^1\right)$.[2]

Our identification theorem is constructive. We iteratively find all moments of β, and thus identify the density f^0 within the class of densities that are uniquely determined by all of their moments. This class is the class of densities that satisfies Carleman's condition, which we review below. Our proof strategy is not unique to the logit: it could be applied to identify the density of heterogeneity in many differentiable economic models. We outline the main theorem using generic notation and verify its main condition for the multinomial logit model.

Showing that the density f^0 is nonparametrically identified is a necessary component for any consistency proof for a nonparametric estimator of f. Indeed, we introduce a computationally simple, nonparametric sieve estimator for f^0 in Bajari, Fox, Kim and Ryan (2009) for general mixtures models. This identification theorem therefore completes our proof of consistency for the estimator of the random coefficients logit in Bajari et al. Identification does not rule out that the operator Q_j has a discontinuous inverse and that the estimation problem is ill-posed. Hence, we use a sieve estimator to gain consistency under a potential ill-posed inverse problem. Alternative nonparametric estimators include the Bayesian MCMC estimator in Rossi, Allenby, and McCulloch

[2] Our identification approach applies to differentiable, and hence continuous, functions $\Pr(j \mid x;\, f)$. Hence, if there exists an x where $\Pr\left(j \mid x;\, f^0\right) \neq \Pr\left(j \mid x;\, f^1\right)$, by continuity there will exist a continuum of x where $\Pr\left(j \mid x;\, f^0\right) \neq \Pr\left(j \mid x;\, f^1\right)$. If x has positive support on an open set in this continuum, there will exist a positive probability of covariates where $\Pr\left(j \mid x;\, f^0\right) \neq \Pr\left(j \mid x;\, f^1\right)$. This positive probability of x's is necessary to prove the consistency of extremum estimators.

(2005) and the EM algorithm used in Train (2008). Neither work discusses consistency, ill-posedness, or identification.

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 density f^0. One possibility was that the normality assumptions typically imposed on f^0 were crucial to identification: without restricting attention to a particular parametric functional form, two f's would indeed solve (2) for all x, even with data on a continuum of x. We show that indeed the random coefficients logit model is identified, which provides backing to its immensely large use in the empirical industrial organization, marketing, transportation, environmental and engineering literatures.

2 Previous and Subsequent Literature

Previously, Ichimura and Thompson (1997) studied the case of binary choice: one inside good ($J = 1$) and one outside good. The binary-choice restriction makes their method inapplicable for most empirical applications to demand analysis. Ichimura and Thompson identify the CDF of, in our notation, $(\beta, \epsilon_1 - \epsilon_0)$. They use a theorem due to Cramér and Wold (1936) and do not exploit the structure of the extreme value assumptions on the ϵ_1 and ϵ_0. Consequently, they need stronger assumptions: a monotonicity assumption (sign restriction) on one of the K components of β ($\beta_k > 0$ for all consumers) and a full support assumption for all K elements of x_1. In contrast, we need only local variation in one x_j within an infinitesimally small open set. Gautier and Kitamura (2008) provide a computationally simpler estimator and some alternative identification arguments (the results are equivalent) for the same binary choice model as Ichimura and Thompson.

The identification of the logit mixtures model occurs by varying the linear index $x_j'\beta$ around a neighborhood of 0. This is a very local form of identification, and is much weaker on the data than identification arguments that rely on identification at infinity, such as Lewbel (2000). On the other hand, Lewbel uses a large-support "special regressor" to avoid our assumption of the independence of x and ϵ, does not require that all elements of x_j be continuous, and does not use the logit, so the two sets of assumptions are non-nested. Lewbel does not identify the distribution of random coefficients, but the centrality parameters $E[\beta]$ and the distribution M of the remaining errors, $M\left(x_j'\beta - x_j'E[\beta] + \epsilon_j \mid x\right)$, which is not enough for some structural uses of demand systems, as explained below in the Berry and Haile discussion.

Subsequent to the circulation of this theorem, Berry and Haile (2008) and Fox and Gandhi (2009) introduced identification arguments for multinomial choice models without the Type I extreme value distribution or additive errors. Like the analysis of binary choice in Ichimura and Thompson

(1997), both Berry and Haile and Fox and Gandhi need a monotonicity assumption on one of the K components of β ($\beta_k > 0$ for all consumers) and (for full identification) a full support assumption on the corresponding kth component $x_{k,j}$, for all choices $j \in J$. By exploiting the functional form assumption on the ϵ_j's, we do not need extreme values of covariates to induce switching behaviour for consumers with very high values of ϵ_j. Berry and Haile identify only the conditional-on-x distribution of utility values $G(u_0, u_1, \ldots, u_J \mid x)$ 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 $H\left(\Delta u \mid x^1, x^2\right)$, where

$$\Delta u = \max_{j \in J \cup \{0\}} u_j\left(x^1\right) - \max_{j \in J \cup \{0\}} u_j\left(x^2\right),$$

where $u_j\left(x^1\right)$ is just the realized utility value (1) for $x^1 = \left(x_1^1, \ldots, x_J^1\right)$. 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 $D(u_1(x), \ldots, u_J(x))$, where $u_j(x)$ is a complete function that describes utility values for choice j at all x. Again, like Berry and Haile, Fox and Gandhi rely on monotonicity and large support assumptions on a single regressor, if the true model has additive errors ϵ_j in it. Fox and Gandhi work in the class of multinomial distributions with unknown numbers and identities of support points, which is non-nested with our class of distributions, those that admit a density satisfying Carleman's condition.[3]

Our paper focuses on continuous covariates in x. All arguments can be made conditional on the values of discrete covariates, but no paper has explored identifying a distribution of random coefficients on discrete covariates in a discrete choice setting.

3 Main Theorem

When stating the main theorem, we shall consider a more general model which includes the random coefficients logit as a special case. In the next section we verify the key condition for the multinomial logit. The econometrician observes covariates x and the probability of some binary outcome, $P(x)$. For a model with a more complex outcome (including a continuous outcome y), we can always consider whether some event ($y < \frac{1}{2}$ say) happened or did not happen. $P(x)$ is the probability of the event happening. x is independent of β. Let $g(x, \beta)$ be the probability of an agent with

[3] Lewbel, Berry and Haile and Fox and Gandhi all discuss using instrumental variables for identification when some regressors are not independent of unobservables. We do not discuss endogenous regressors here, in concert with many empirical applications of the multinomial logit model.

characteristics β taking the action. Our goal is to identify the density function f in the equation

$$P(x) = \int g(x, \beta) f(\beta) d\beta. \tag{3}$$

Identification means that a unique f solves this equation for all x.[4]

Here we propose to identify the density of f by finding its moments when g is differentiable and satisfies the single-index condition $g(x, \beta) = g(x'\beta)$. A probability measure 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.

Assumption 3.1

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

The Carleman condition gives uniqueness for distributions with unrestricted support. If the support of β is known and compact, uniqueness follows without the Carleman condition. The component function $g(x'\beta)$ does not have to be a distribution function. We mainly require that $g(x'\beta)$ be continuously differentiable. We do heavily exploit the linear index $x'\beta$.

Assumption 3.2

- *$g(x'\beta) \in C^\infty$ (infinitely continuously differentiable) in a neighborhood of $x = 0$.*
- *$g^{(l)}(0)$ is nonzero and finite for all $l \geq 1$ where $g^{(l)}(\cdot)$ denotes the lth derivative of $g(\cdot)$.*

Assumption 3.2 restricts the class of $g(x'\beta)$. Some classes of functions satisfy the condition but others do not. For example, if $g(w) = C \cdot \exp(w)$, then Assumption 3.2 is trivially satisfied, because $g^{(l)}(0) = C$ for all l. If $g(x'\beta)$ 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 density f up to the vth moment, where v is the order of the polynomial function. Because of our focus on differentiability, we require covariates with continuous support, but not at all wide support.

Assumption 3.3

- *The covariates in the vector x take on support in an open set containing $x = 0$.*

[4] This is the definition used in the statistics literature, see Teicher (1963).

We observe $P(x)$ in the population data and know the function g. We wish to identify the density f. The general identification argument can be illustrated for the special case where $K = 2$ and so $x'\beta = x_1\beta_1 + x_2\beta_2$. At $x_1 = x_2 = 0$,

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

where β_1 arises from the chain rule and the expression identifies the mean of β_1, because $P(x)$ is data and $g^{(1)}(0)$ is a known constant that does not depend on β.[5] Likewise, $\left.\frac{\partial P(x)}{\partial x_2}\right|_{x=0} / g^{(1)}(0)$ equals $E[\beta_2]$, $\left.\frac{\partial^2 P(x)}{\partial x_1 \partial x_2}\right|_{x=0} / g^{(2)}(0)$ equals $E[\beta_1\beta_2]$, and $\left.\frac{\partial^2 P(x)}{\partial^2 x_1}\right|_{x=0} / g^{(2)}(0)$ equals $E\left[\beta_1^2\right]$. Additional derivatives will identify the other moments of $\beta = (\beta_1, \beta_2)$.

Theorem 3.1

- *Suppose Assumptions 3.1, 3.2 and 3.3 hold. Then the true f^0 is identified.*
- *Assume the lth derivative of $g(z)$ is nonzero when evaluated evaluated at $z = 0$. Then under Assumption 3.3, all moments of order l (including cross moments) of the elements of the vector β are identified.*

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

The problem of identifying a distribution uniquely from the first L moments of the corresponding random variable is known as the determinacy (unique solution) of the truncated Hamburger moment problem (Akhiezer 1965, Krein and Nudel'man 1973). Truncated moment problems are a well studied topic in probability theory. A key tool is a Hankel matrix, which is formed from the first L moments. If the Hankel matrix has a zero determinant, a unique distribution has these particular L moments. Extensions of these results exist for multidimensional random variables (Akhiezer 1965). What is important here is that results on the truncated moments problem exist and are not related to the type of economic model in which the unobserved heterogeneity enters.

[5] Assumption 3.2 allows us to exchange differentiation and integration, via Leibniz's integral rule.

4 Identification of The Logit Model

We can fit the random coefficients logit model into the mixtures framework by defining the logit choice probabilities for some particular choice j as

$$g(x, \beta) = \frac{\exp\left(x_j'\beta\right)}{1 + \sum_{h=1}^{J} \exp\left(x_h'\beta\right)}.$$

To highlight our main result, we state as a theorem that the logit model is identified. Surprisingly, identification in the random coefficients logit model has never been proved despite its 30 years of use.

Theorem 4.1

- *Let Assumptions 3.1 and 3.3 hold. Let $g(\cdot)$ be the random coefficients logit model with $J \geq 2$ inside goods and one outside good. Then the true f^0 is identified.*

The proof in the appendix uses Theorem 3.1; the proof consists largely of verifying Assumption 3.2. The key idea in the proof is the use of the "rational zero test", which allows us to focus on the integer solutions to polynomials that arise in the expressions for the derivatives of the logit choice probabilities.

For the case of one inside good ($J = 1$), algebra (for a known order of derivatives) shows that $g^{(l)}(0) \neq 0$ when l is odd and $g^{(l)}(0) = 0$ when l is even. The zero derivatives mean that we will need to impose that the true density of β generates statistically independent random variables. In other words, we show the logit mixtures model with $f(\beta) = \prod_{k=1}^{K} f_k(\beta_k)$ (an independent multivariate distribution) that also satisfies Assumptions 3.1 and 3.3 is identified. We identify the odd moments of β using Theorem 3.1, but because $f(\beta) = \prod_{k=1}^{K} f_k(\beta_k)$, we also identify any even moments. To see this for an example, from Theorem 3.1 we can obtain two odd moments such as $E\left[\beta_1\beta_2^2\right]$ and $E[\beta_1]$. As β_1 and β_2 are independent, we also obtain the even moment $E\left[\beta_2^2\right] = E\left[\beta_1\beta_2^2\right] / E[\beta_1]$.

5 Conclusions

The random coefficients logit model has been used in empirical studies for over 30 years. We are the first to show that the density of random coefficients is nonparametrically identified. This allows complete proofs for the consistency of nonparametric estimators of the density of random coefficients. We also confirm to empirical researchers that identification of the density of preferences relies on variation in covariates and not only on functional form assumptions for the density.

Compared to previous and subsequent identification results in the literature for binary and multinomial choice, we exploit the type I extreme value distribution on the additive errors. Thus, we remove the need to consider the monotonicity and large support assumptions needed in the literature.

A Proofs of the Theorems

A.1 Proof of Theorem 3.1

First we introduce some notation for gradients of arbitrary order, which we need because $f(\beta)$ has a vector of K arguments, β. Let w be a vector of length W. For a function $h(w)$, we denote the $1 \times K^v$ block vector of vth order derivatives as $\nabla^v h(w)$. $\nabla^v h(w)$ is defined recursively so that the kth block of $\nabla^v h(w)$ is the $1 \times W$ vector $h_k^v(w) = \partial h_k^{v-1}(\theta)/\partial w'$, where h_k^{v-1} is the kth element of $\nabla^{v-1} h(w)$. Using a Kronecker product $\otimes$, we can write $\nabla^v h(w) = \frac{\partial^v h(w)}{\underbrace{\partial w' \otimes \partial w' \otimes \ldots \otimes \partial w'}_{v \text{ Kronecker product of } \partial w'}}$.

Take the derivatives with respect to the covariates x of both sides of $P(x) = \int g(x'\beta) f(\beta) d\beta$ and evaluate the derivatives at $x = 0$. By Assumption 3.2, for any $v = 1, 2, \ldots$ and the chain rule repeatedly applied to the linear index $x'\beta$,

$$\begin{aligned} \nabla^v P(x)\big|_{x=0} &= \int g^{(v)}\left(x'\beta\right)\Big|_{x=0} \left\{\beta' \otimes \beta' \otimes \cdots \otimes \beta'\right\} f(\beta)\, d\beta \qquad (4) \\ &= g^{(v)}(0) \int \left\{\beta' \otimes \beta' \otimes \cdots \otimes \beta'\right\} f(\beta)\, d\beta. \end{aligned}$$

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 $v = 1, 2, \ldots.,$ we obtain the vth moments of f for all $v \geq 1$. Now by Assumption 3.1, f satisfies the Carleman condition. Therefore, f is identified since a probability measure satisfying the Carleman condition is uniquely determined by its moments.

A.2 Proof of Theorem 4.1

Identification arises from identifying all moments, as in Theorem 3.1. The main condition to verify is Assumption 3.2: all derivatives are nonzero when evaluated at 0. Let $x_h = 0$ for all $h \neq j$. With one outside good and J inside goods, the choice probability of alternative j given β is

$$g_j\left(x_1'\beta,\ldots,x_j'\beta,\ldots,x_J'\beta\right)=g_j\left(0,\ldots,x_j'\beta,\ldots,0\right)=\frac{\exp\left(x_j'\beta\right)}{1+(J-1)+\exp\left(x_j'\beta\right)}=\frac{\exp\left(x_j'\beta\right)}{J+\exp\left(x_j'\beta\right)}.$$

Define $g_J(a)=\frac{e^a}{J+e^a}$ and let D_a^p denote the derivative operator of order p with respect to a. We wish to show

$$D_a^p g_J(a)\big|_{a=0}\neq 0 \text{ for all integer } J\geq 2 \text{ and for all } p.$$

We obtain

$$\begin{aligned}
D_a g_J(a) &= \frac{1}{(J+e^a)^2}Je^a,\ D_a^2 g_J(a)=\frac{1}{(J+e^a)^3}\left(J^2e^a-Je^{2a}\right)\\
D_a^3 g_J(a) &= \frac{1}{(J+e^a)^4}\left(J^3e^a-4J^2e^{2a}+Je^{3a}\right)\\
D_a^4 g_J(a) &= \frac{1}{(J+e^a)^5}\left(J^4e^a-11J^3e^{2a}+11J^2e^{3a}-Je^{4a}\right)\\
D_a^5 g_J(a) &= \frac{1}{(J+e^a)^6}\left(J^5e^a-26J^4e^{2a}+66J^3e^{3a}-26J^2e^{4a}+Je^{5a}\right)\\
&\vdots
\end{aligned}$$

For $p\geq 3$, now we denote the $(p-1)$th derivative as

$$D_a^{p-1}g_J(a)=\frac{1}{(J+e^a)^p}\sum_{j=1}^{p-1}\theta_{p-j}^{(p)}J^{p-j}e^{ja}.$$

Then, we can write the p-th derivative as

$$
\begin{aligned}
& D_a^p g_J(a) \\
= \quad & \left[\frac{1}{(J+e^a)^{p+1}} \sum_{j=1}^{p} \theta_{p+1-j}^{(p+1)} J^{p+1-j} e^{ja} \right] \qquad (5) \\
= \quad & D_a D_a^{p-1} g_J(a) \\
= \quad & D_a \left[\frac{1}{(J+e^a)^p} \sum_{j=1}^{p-1} \theta_{p-j}^{(p)} J^{p-j} e^{ja} \right] \\
= \quad & \frac{1}{(J+e^a)^p} \sum_{j=1}^{p-1} j\theta_{p-j}^{(p)} J^{p-j} e^{ja} - \frac{1}{(J+e^a)^{p+1}} p \sum_{j=1}^{p-1} \theta_{p-j}^{(p)} J^{p-j} e^{(j+1)a} \\
= \quad & \frac{1}{(J+e^a)^{p+1}} \left((J+e^a) \sum_{j=1}^{p-1} j\theta_{p-j}^{(p)} J^{p-j} e^{ja} - \sum_{j=1}^{p-1} p\theta_{p-j}^{(p)} J^{p-j} e^{(j+1)a} \right) \\
= \quad & \frac{1}{(J+e^a)^{p+1}} \left(\sum_{j=1}^{p-1} j\theta_{p-j}^{(p)} J^{p+1-j} e^{ja} + \sum_{j=1}^{p-1} j\theta_{p-j}^{(p)} J^{p-j} e^{(j+1)a} - \sum_{j=1}^{p-1} p\theta_{p-j}^{(p)} J^{p-j} e^{(j+1)a} \right) \\
= \quad & \frac{1}{(J+e^a)^{p+1}} \left(\begin{array}{c} \theta_{p-1}^{(p)} J^p e^a + \sum_{j'=2}^{p-1} j'\theta_{p-j'}^{(p)} J^{p+1-j'} e^{j'a} + \sum_{j=1}^{p-1} j\theta_{p-j}^{(p)} J^{p-j} e^{(j+1)a} \\ - \sum_{j=1}^{p-1} p\theta_{p-j}^{(p)} J^{p-j} e^{(j+1)a} \end{array} \right) \qquad (6) \\
= \quad & \frac{1}{(J+e^a)^{p+1}} \left(\begin{array}{c} \theta_{p-1}^{(p)} J^p e^a + \sum_{j=1}^{p-2} (j+1)\theta_{p-j-1}^{(p)} J^{p-j} e^{(j+1)a} + \sum_{j=1}^{p-1} j\theta_{p-j}^{(p)} J^{p-j} e^{(j+1)a} \\ - \sum_{j=1}^{p-1} p\theta_{p-j}^{(p)} J^{p-j} e^{(j+1)a} \end{array} \right) \qquad (7) \\
= \quad & \frac{1}{(J+e^a)^{p+1}} \left(\theta_{p-1}^{(p)} J^p e^a + \sum_{j=1}^{p-2} \{(j+1)\theta_{p-j-1}^{(p)} + j\theta_{p-j}^{(p)} - p\theta_{p-j}^{(p)}\} J^{p-j} e^{(j+1)a} - \theta_1^{(p)} J^1 e^{pa} \right) \qquad (8)
\end{aligned}
$$

where in (6) and (7), we take out the first element in the first sum and change the index j' to $j+1$. (8) is obtained by rearranging terms and collecting coefficients on $J^{p-j}e^{(j+1)a}$ for $j = 1$ to $p-2$.

To fix the undetermined coefficients $\theta_{p-j}^{(p)}$'s, we compare the coefficients from (5) and (8) and obtain

$$\begin{aligned}
\sum_{j=1}^{p} \theta_{p+1-j}^{(p+1)} J^{p+1-j} e^{ja} &= \theta_p^{(p+1)} J^p e^a + \sum_{j=2}^{p-1} \theta_{p+1-j}^{(p+1)} J^{p+1-j} e^{ja} + \theta_1^{(p+1)} J^1 e^{pa} \\
&= \theta_p^{(p+1)} J^p e^a + \sum_{j=1}^{p-2} \theta_{p-j}^{(p+1)} J^{p-j} e^{(j+1)a} + \theta_1^{(p+1)} J^1 e^{pa} \\
&= \theta_{p-1}^{(p)} J^p e^a + \sum_{j=1}^{p-2} \left\{ (j+1)\, \theta_{p-j-1}^{(p)} + j\theta_{p-j}^{(p)} - p\theta_{p-j}^{(p)} \right\} J^{p-j} e^{(j+1)a} - \theta_1^{(p)} J^1 e^{pa}.
\end{aligned}$$

We find

$$\begin{aligned}
\theta_p^{(p+1)} &= \theta_{p-1}^{(p)} && (9) \\
\theta_{p-j}^{(p+1)} &= (j+1)\, \theta_{p-j-1}^{(p)} + j\theta_{p-j}^{(p)} - p\theta_{p-j}^{(p)} \text{ for } p \geq 3 \\
\theta_1^{(p+1)} &= -\theta_1^{(p)}. && (10)
\end{aligned}$$

This system generates the coefficients for all $p \geq 1$. For the initial value, we obtain $\theta_1^{(2)} = 1$. When $p = 2$, we find

$$\begin{aligned}
\theta_2^{(3)} &= \theta_1^{(2)} = 1 \\
\theta_1^{(3)} &= -\theta_1^{(2)} = -1
\end{aligned}$$

and when $p = 3$, we find

$$\begin{aligned}
\theta_3^{(4)} &= \theta_2^{(3)} = 1 \\
\theta_2^{(4)} &= 2\theta_1^{(3)} + \theta_2^{(3)} - 3\theta_2^{(3)} = -4 \\
\theta_1^{(4)} &= -\theta_1^{(3)} = 1.
\end{aligned}$$

Now we examine whether $D_a^p g_J(a)|_{a=0}$ can take the value of zero for some p and some J. For this purpose, we evaluate the derivatives at $a = 0$ and obtain equations with respect to J for the pth order derivative as

$$D_a^p g_J(a)|_{a=0} = \frac{1}{J^{p+1}} \sum_{j=1}^{p} \theta_{p+1-j}^{(p+1)} J^{p+1-j} = 0$$

for all $p \geq 1$. This is equivalent to solving

$$\sum_{j=1}^{p} \theta_{p+1-j}^{(p+1)} J^{p-j} = 0. \tag{11}$$

Now note that the coefficient on J^{p-1} (the highest order term in the equation) in (11) is equal to

$$\theta_p^{(p+1)} = 1$$

for all p. Also note that the constant term (the coefficient on J^0 in (11)), $\theta_1^{(p+1)}$, is equal to 1 when p is odd and is equal to -1 when p is even. By the well-known "rational zero test", this implies that the only possible positive integer solution in (11) is $J = 1$. A positive integer greater than 1 cannot be the solution of (11) for any p. This concludes our claim.

References

[1] Akhiezer, N.I. (1965), *The classical moment problem and some related questions in analysis*, Oliver & Boyd.

[2] Bajari, P, JT Fox, KI Kim, and S Ryan (2009), "A Simple Nonparametric Estimator for the Distribution of Random Coefficients in Discrete Choice Models", working paper.

[3] Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Price in Market Equilibrium", *Econometrica* (63), July 1995.

[4] Berry, S and Haile, P. (2008), "Nonparametric Identification of Multinomial Choice Models with Heterogeneous Consumers and Endogeneity", working paper.

[5] Boyd, JH and RE Mellman (1980), "Effect of Fuel Economy Standards on the US Automotive Market", *Transportation Research.* 14(5), 367–378.

[6] Cardell, NS and FC Dunbar (1980), "Measuring the societal impacts of automobile downsizing", *Transportation Research.* 14(5), 423–434.

[7] Cramér, H. and H. Wold (1936), "Some Theorems on Distribution Functions", *Journal of the London Mathematical Society*, s1-11(4), 290–294.

[8] Fox, Jeremy T. and Amit Gandhi (2009), "Identifying Heterogeneity in Economic Choice and Selection Models Using Mixtures", working paper.

[9] Gautier, Eric and Kitamura, Yuichi. (2008), "Nonparametric Estimation in Random Coefficients Binary Choice Models", working paper.

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

[11] 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.

[12] Krein, M.G. and A.A. Nudel'man (1973, translation 1977), *The Markov moment problem and extremal problems: ideas and problems of P. L. Čebyšev and A. A. Markov and their further*

development. Translations of mathematical monographs v. 50, American Mathematical Society, Providence.

[13] Lewbel, Arthur. (2000) "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables", *Journal of Econometrics*, 97, 1, 145-177.

[14] McFadden, Daniel and Kenneth Train (2000), "Mixed MNL models for discrete response", *Journal of Applied Econometrics*, 15(5): 447–470.

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

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

[17] Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2005), *Bayesian Statistics and Marketing.* West Sussex: John Willy & Sons.

[18] Shohat, J.A. and Tamarkin, J.D. (1943), *The Problem of Moments*, American Mathematics Society, Providence, RI.

[19] Teicher, H. (1963), "Identifiability of Finite Mixtures", *Annals of Mathematical Statistics*, 34, 1265-1269.

[20] Train, Kenneth. 2003. *Discrete choice methods with simulation.* Cambridge.

[21] Train, Kenneth. 2008. "EM Algorithms for Nonparametric Estimation of Mixing Distributions." *Journal of Choice Modeling,* 1, 1, 40–69.