Identification of Random Coefficient Latent Utility Models[*]

Roy Allen

Department of Economics

University of Western Ontario

rallen46@uwo.ca

John Rehbeck

Department of Economics

The Ohio State University

rehbeck.7@osu.edu

March 2, 2020

Abstract

This paper provides nonparametric identification results for random coefficient distributions in perturbed utility models. We cover discrete and continuous choice models. We establish identification using variation in mean quantities, and the results apply when an analyst observes aggregate demands but not whether goods are chosen together. We require exclusion restrictions and independence between random slope coefficients and random intercepts. We do not require regressors to have large supports or parametric assumptions.

---

[*]Any remaining errors are our own.

1 Introduction

Latent utility models with linear random coefficients have been extensively used. They have a long history in discrete choice,[1] and have become increasingly popular due to computational advances (see e.g. Train [2009]). They now form the core demand system of most applied work involving demand for differentiated products following Berry et al. [1995]. Progress has been made on identification of these models in discrete choice, but gaps remain, even in semiparametric settings. For example, nonparametric identification of the distribution of random coefficients in the random coefficients nested logit model has not been established without unbounded regressors.[2] More broadly, there is growing interest in models that allow complementarity, but even less is known about identification of random coefficients in these models.[3]

The main contribution of this paper establishes nonparametric identification for the moments of random coefficients in a general class of latent utility models. The framework applies to discrete and continuous choices. As a special case, we establish identification for a bundles model with limited consideration of either alternatives or characteristics (Example 2). Identification only depends on the average structural function [Blundell and Powell, 2003]. Thus the results can be applied when one observes the average demands of individuals without observing whether goods are chosen together. Leveraging the main result, we can identify the distribution of random coefficients when it is characterized by its moments (e.g. normal distributions). Specialized to discrete choice, the main contribution is new since it does not require any regressor to be unbounded.

Two key ingredients let us get traction for identification. First, we assume independence between random slopes and random intercepts. This is a standard assumption in the widely-used random coefficients logit model. We use this assumption to integrate out the random intercepts. This smooths out demand when conditioning on

---

[1]Heckman [2001] attributes the first use to Domenich and McFadden [1975] in economics.

[2]See Nevo [2000], p. 524-526. il Kim [2014] has established identification in the special case where intercept location coefficients are 0.

[3]Recent work includes Gentzkow [2007], McFadden and Fosgerau [2012], Fosgerau et al. [2019], Allen and Rehbeck [2019a], Ershov et al. [2018], Monardo [2019], Iaria and Wang [2019], and Wang [2020]. This work is an outgrowth of the discrete choice additive random utility model [McFadden, 1981] and differs from classic continuous demand systems (e.g. Deaton and Muellbauer [1980]) by focusing on characteristic variation rather than variation in a budget constraint.

non-intercept components of the utility function (regressors and random coefficients). This also allows us to treat discrete and continuous choice models in a common framework. Second, we exploit theoretical restrictions since choices arise from optimization. Without using this structure, the model would resemble general random coefficients index models studied in Fox et al. [2012] and Lewbel and Pendakur [2017]. While they assume a function governing the mapping from indices to choices is *known*, we do not.[4]

The fundamental shape restriction we exploit is that, after integrating out random intercepts, integrated mean choices are the derivative of a convex function. This follows from an application of the envelope theorem. Similar tools from convex analysis have also been used for identification of hedonic models [Ekeland et al., 2002, 2004, Heckman et al., 2010, Chernozhukov et al., 2019b], matching [Galichon and Salanié, 2015], dynamic discrete choice [Chiong et al., 2016], discrete choice panel models [Shi et al., 2018], and perturbed models with additively separable heterogeneity [Allen and Rehbeck, 2019a], among others.[5]

By exploiting the envelope theorem, we can treat several models in a common framework. There is little work on identification of optimizing models with linear random coefficients outside of discrete choice. Exceptions include Dunker et al. [2018] for discrete games and Dunker et al. [2017] and Iaria and Wang [2019] for a random coefficients version of the Gentzkow [2007] discrete bundles model. We differ by requiring identification of only the average structural function ("mean demands"), without needing to observe the frequency with which goods are chosen together.[6] Identification with linear random coefficients has also been established in settings without assuming an optimizing model. See for example the simultaneous equations analysis in Masten [2017] and references therein.

Identification of linear random coefficients has been extensively studied in discrete choice. Despite this, nonparametric identification has only been established either

---

[4]In discrete choice, assuming this function is known translates to the distribution of random intercepts being known (e.g. logit). Lewbel and Pendakur [2017] show that one can drop the assumption that this mapping is known in some settings if one imposes an additional additive separability assumption.

[5]Matzkin [1994] reviews other identification results using shape restrictions motivated by economic theory. See also work on optimal transport, as in Galichon [2018].

[6]Wang [2020] also works with the average structural function but does not identify the distribution of random coefficients.

requiring a regressor with large support or assuming the distribution of random intercepts is either known or parametric. One reason we do not require a large support assumption is that we focus on identification of the distribution of random coefficients without identifying random intercepts. Papers that make use of large support regressors with heterogeneity that is not additively separable include [Ichimura and Thompson, 1998], Berry and Haile [2009], Briesch et al. [2010], Gautier and Kitamura [2013], Fox and Gandhi [2016], Dunker et al. [2017], and Fox [2017].[7] Several of these papers additionally assume large support regressors *also* have the same coefficient across goods. In contrast, Fox et al. [2012] and il Kim [2014] do not assume large support or a homogeneous regressor, but assume the distribution of the random intercept is known (e.g. logit). Chernozhukov et al. [2019a] discusses identification of ratios of certain moments of the distribution of random coefficients without requiring large support, but do not provide conditions under which the full distribution is identified.

The remainder of the paper proceeds as follows. Section 2 provides details on the class of latent utility models we study and examples of behavior that this covers. Section 3 provides the main result, which identifies arbitrary order moments of random coefficients and shows that a single independence and scale assumption can be used to identify all other moments. Section 4 discusses how to recover different welfare objects and perform counterfactuals. Finally, Section 5 discusses relations to some existing papers, shows how the results can be taken to settings with non-linear random coefficients, and discusses some testable properties of the framework.

2 Setup

This paper studies the *random coefficients perturbed utility model*, in which optimizing choices satisfy

$$Y(X, \beta, \varepsilon) \in \underset{y \in B}{\operatorname{argmax}} \sum_{k=1}^{K} y_k(\beta_k' X_k) + D(y, \varepsilon). \tag{1}$$

---

[7]Exceptions include Kashaev [2018] and Matzkin [2019], but neither paper studies nonparametric identification for the distribution of linear random coefficients.

McFadden and Fosgerau [2012] and Allen and Rehbeck [2019a] have studied related frameworks without random coefficients. We interpret $Y(\cdot)$ as the quantity vector for K different goods. The vector $X_k = (X_{k,1}, \ldots, X_{k,d_k})'$ denotes observable shifters of the desirability of good k, and $\beta_k = (\beta_{k,1}, \ldots, \beta_{k,d_k})'$ denotes random coefficients on these shifters, which may be good-specific. The index $\beta_k' X_k$ shifts the marginal utility of good k. We collect $X = (X_1', \ldots, X_K')'$ and $\beta = (\beta_1', \ldots, \beta_K')'$. The term $D(y, \varepsilon)$ is a disturbance that depends on unobservables ε of unrestricted dimension. When $D(y, \varepsilon) = \sum_{k=1}^K \varepsilon_k y_k$, ε_k can be interpreted as a random intercept for the desirability of the k-th good. In general, we refer to $D(y, \varepsilon)$ as the random intercept. The set $B \subseteq \mathbb{R}^K$ is a feasibility set. This is introduced purely for exposition, since $D(y, \varepsilon)$ can be $-\infty$ which allows random feasibility sets.

The focus of this paper is on identification of moments of the distribution of β. Our results do not require specification of the budget B, the disturbance D, or the distribution of over ε. For concreteness, we provide some examples.

Example 1 (Discrete Choice). *Consider a discrete choice models with latent utility for good k of the form*

$$v_k = \beta_k' X_k + \varepsilon_k.$$

When β is random, this is a linear random coefficients model as studied in Hausman and Wise [1978], Boyd and Mellman [1980], Cardell and Dunbar [1980], among many others. This fits into the setup of (1) by setting $D(y, \varepsilon) = \sum_{k=1}^K y_k \varepsilon_k$, $B = \{y \in \mathbb{R}^K \mid \sum_{k=1}^K y_k = 1, y_k \geqslant 0\}$ the probability simplex,[8] and letting $Y(X, \beta, \varepsilon) \in \{0, 1\}^K$ be a vector of indicators denoting which good is chosen. In many applications, an "outside good" is set to have a utility of 0. This can be mapped to our setup by replacing the budget with $B = \{y \in \mathbb{R}^K \mid \sum_{k=1}^K y_k \leqslant 1, y_k \geqslant 0\}$; this allows $Y(X, \beta, \varepsilon) = (0, \ldots, 0) \in B$, which can be interpreted as the choice of the outside option.

We also cover what is sometimes called the perturbed representation of choice, which can model market shares or individuals who like variety. For example, Anderson et al. [1988] show logit models are related to the maximization problem

$$\max_{y \in \Delta} \sum_{k=1}^K y_k (\beta_k' X_k) + \sum_{k=1}^K y_k \log(y_k),$$

---

[8]This allows the agent to randomize when there are utility ties.

with Δ the probability simplex. Hofbauer and Sandholm [2002] show by replacing the additive entropy term with a general disturbance, the setup covers all discrete choice additive random models once random intercepts are integrated out. Fudenberg et al. [2015] study a model in which the disturbance is additively separable. Fosgerau et al. [2019] and Allen and Rehbeck [2019b] show how to model complementarity with the perturbed utility representation.

Example 2 (Bundles with Limited Consideration). *Gentzkow [2007] presents a model of choice of bundles involving online and print news. The model involves multiple goods and individuals can choose more than one good at the same time. A random coefficients version of the model has been studied in Dunker et al. [2017]. Let $v_{j,k}$ denote the utility associated with quantity j of the first good, and quantity k of the second good. Specify utilities*

$$v_{0,0} = 0$$
$$v_{1,0} = \beta_1' X_1 + \varepsilon_{1,0}$$
$$v_{0,1} = \beta_2' X_2 + \varepsilon_{0,1}$$
$$v_{1,1} = v_{1,0} + v_{0,1} + \varepsilon_{1,1}.$$

Here, $\varepsilon_{1,1}$ denotes a utility boost or loss from purchasing both goods relative to the sum of their individual utility. It describes complementarity/substitutability between the goods. For each quantity vector $\vec{y} = (y_1, y_2)$, set the utility as

$$\sum_{k=1}^{2} y_k (\beta_k' X_k) + (y_1 \varepsilon_{1,0} + y_2 \varepsilon_{0,1} + y_1 y_2 \varepsilon_{1,1}),$$

and let the budget be $B = \{0,1\}^2$. Then the optimizing quantity vector $Y(X, \beta, \varepsilon) \in \{0,1\}^2$ fits into the setup of (1).

This can be modified to include latent budgets. One may interpret these as mental "consideration sets" [Eliaz and Spiegler, 2011, Masatlioglu et al., 2012, Manzini and Mariotti, 2014, Aguiar, 2017] or general latent feasibility sets [Manski, 1977, Conlon and Mortimer, 2013, Brady and Rehbeck, 2016]. In addition, we can allow limited consideration of the characteristics of goods.[9]

To model these types of limited attention, consider a version of the bundles model given by

$$Y(X, \beta, \varepsilon) \in \operatorname*{argmax}_{y \in \{0,1\}^2} \sum_{k=1}^{2} y_k(\beta_k' X_k) + D(y, \varepsilon),$$

where

$$D(y, \varepsilon) = \begin{cases} y_1 \varepsilon_{1,0} + y_2 \varepsilon_{0,1} + y_1 y_2 \varepsilon_{1,1} & \textit{if } y \in B(\varepsilon) \\ -\infty & \textit{otherwise} \end{cases}.$$

Here, the set $B(\varepsilon)$ is a latent feasibility set, which could arise because an individual may not consider all goods or the analyst cannot observe when goods are out of stock. Some components of β can be zero with positive probability, reflecting that individuals may not notice or care about certain characteristics.

This setup can be generalized to allow more goods, with some goods continuous and some goods discrete. What is key for our analysis is that the index $\beta_k' X_k$ shifts (only) the marginal utility of good k.

2.1 Average Structural Function and Endogeneity

This paper establishes identification of moments of β using the average structural function [Blundell and Powell, 2003]

$$\overline{Y}(x) = \int Y(x, \beta, \varepsilon) d\tau(\beta, \varepsilon)$$

for some probability measure τ that does not depend on covariates x. We assume that the measure τ satisfies a key independence condition.

Assumption 1 (Slope-Intercept Independence)**.** *The random variables β and ε are independent under the measure τ, and the average structural function is finite.*

While independence between β and ε is restrictive, it is a standard assumption in applications of the random coefficients logit model in discrete choice. It has been exploited for identification in [Fox et al., 2012] and [Chernozhukov et al., 2019a].[10]

---

[10]However, independence is not imposed in some papers studying identification. For example, Ichimura and Thompson [1998] or Gautier and Kitamura [2013] do not impose independence of the slope and intercept.

With this assumption, we can write

$$\overline{Y}(x) = \int \int Y(x, \beta, \varepsilon) d\mu(\varepsilon) d\nu(\beta)$$

for some probability measures μ and ν. Technically, full independence is not needed as long as we can factor the average structural function in this way.

For an example of an average structural function, suppose $(Y, X, \beta, \varepsilon)$ are random variables that satisfy $Y = Y(X, \beta, \varepsilon)$ almost surely. Moreover, assume X, β, and ε are all independent. In addition to independence, suppose a continuous version of the conditional mean of Y given X exists. Then

$$\mathbb{E}[Y \mid X = x] = \overline{Y}(x) = \int \int Y(x, \beta, \varepsilon) d\mu(\varepsilon) d\nu(\beta)$$

for x in the support of X,[11] where μ is the marginal distribution of ε and ν is the marginal distribution of β.

The results in this paper apply to general average structural functions $\overline{Y}(x)$, not only the conditional mean. Thus, while slope-intercept independence is important for our results, independence between X and (β, ε) is not. Therefore, the results in this paper are relevant for settings with endogeneity.

The goal of this paper is not to provide a new method to identify the average structural function, but rather to use the function to identify other features of a utility maximizing model. There is a large literature on identifying structural functions. Blundell and Powell [2003] describe how to use control functions to identify the average structural function $\overline{Y}(x)$. Altonji and Matzkin [2005] identify derivatives of the average structural function using certain conditional independence or symmetry conditions. Berry [1994], Berry et al. [1995], Newey and Powell [2003], Berry and Haile [2014], and Dunker et al. [2017] among others use instrumental variables to identify an average structural function from aggregate data.[12]

---

[11]Recall that the support of X is the smallest closed set S such that $P(X \in S) = 1$.

[12]A key step to apply these methods is injectivity in a market-level observable to a vector of unobservable endogenous vectors, usually denoted ξ. See Allen [2019] or Lemma 3 in Allen and Rehbeck [2019a] for injectivity results that cover the present model when the utility index for good k is $\beta_k' x_k + \xi_k$. Related injectivity results have appeared in Galichon and Salanié [2015] and Chiong et al. [2017].

An important feature of the analysis is that only the average structural function is required to be identified over an appropriate region. Thus, the full distribution of $Y(x, \cdot, \cdot)$ induced by the product measure $\mu \times \nu$ over (β, ε) is not necessary for identification. For common discrete choice models the average structural function and the full distribution of $Y(x, \cdot, \cdot)$ contain the same information, but this is not true in general. This is particularly important when combining this analysis with work allowing endogeneity between X and (β, ε). In particular, there are well-understood methods to identify the average structural function in the presence of endogeneity as mentioned earlier. In contrast, less is known about identification of the entire distribution of $Y(x, \cdot, \cdot)$ in the presence of endogeneity.[13]

In addition, requiring only the average structural function implies that the analysis can be applied to settings outside of discrete choice *without* observing whether goods are chosen together. Of course if the full distribution of $Y(x, \cdot, \cdot)$ is identified, then these results apply as well. We recall that this paper does not study identification of the distribution of ε in the original latent utility model (1). However, it is possible to identify the distribution of ε in some cases. For example, Dunker et al. [2017] show how to identify the distribution of random intercepts in a full-consideration random coefficients bundles model, provided the analyst has aggregate data on the frequency with which goods are chosen together.

2.2 Technical Tools

We make use of an aggregation result that first integrates out the distribution of ε.

Lemma 1 (Allen and Rehbeck [2019a]). *Let $Y(\cdot)$ satisfy (1). For any measure μ over ε such that $\int Y(x, \beta, \varepsilon) d\mu(\varepsilon)$ and $\int D(Y(x, \beta, \varepsilon), \varepsilon) d\mu(\varepsilon)$ exist and are finite, it follows that*

$$\int Y(x, \beta, \varepsilon) \mu(d\varepsilon) \in \underset{y \in B}{\operatorname{argmax}} \sum_{k=1}^{K} y_k(\beta_k' x_k) + \overline{D}(y),$$

---

[13]Imbens and Newey [2009] identify average and quantile structural functions with multidimensional heterogeneity in the outcome equation. Torgovitsky [2015] and D'Haultfœuille and Février [2015] identify the entire structural function with one-dimensional unobservable heterogeneity in the outcome equation. A multidimensional counterpart has been studied in Gunsilius [2019]. These papers all identify features of structural functions in the presence of endogeneity.

for $\overline{B}$ the convex hull of B, and $\overline{D}(y) = \sup_{\tilde{Y} \in \mathcal{Y}: \int \tilde{Y}(\varepsilon) d\mu(\varepsilon) = y} \int D\left(\tilde{Y}(\varepsilon), \varepsilon\right) d\mu(\varepsilon),$[14] *where $\mathcal{Y}$ is the set of ε-measurable functions that map to B.*

In addition, the (integrated) indirect utility function

$$V(\beta_1' x_1, \ldots, \beta_K' x_K) = \max_{y \in \overline{B}} \sum_{k=1}^{K} y_k(\beta_k' x_k) + \overline{D}(y),$$

satisfies

$$V(\beta_1' x_1, \ldots, \beta_K' x_K) = \int \left(\max_{y \in B} \sum_{k=1}^{K} y_k(\beta_k' x_k) + D(y, \varepsilon) \right) d\mu(\varepsilon).$$

Note that assuming $Y(\cdot)$ satisfies (1) requires the argmax set to be nonempty. This is a behavioral restriction that imposes sufficient structure for the theorem to go through, and imposes minimal restrictions on D. In particular, D can be $-\infty$ for certain combinations of (y, ε) and need not be continuous. This allows us to treat limited consideration models as in Example 2.

We leverage the aggregation result from Lemma 1 to use calculus-based techniques for identification. To illustrate how aggregation can lead to smoothness, recall that $Y(x, \beta, \varepsilon)$ in discrete choice is a vector of indicators denoting which good is chosen (assuming no ties). Derivatives with respect to x either do not exist at certain points, or are zero and contain little information.

We smooth choices by working with

$$\overline{Y}(x, \beta) := \int Y(x, \beta, \varepsilon) d\mu(\varepsilon)$$

with μ as in Assumption 1. In discrete choice, when ε is integrated out $\overline{Y}(x, \beta)$ can be interpreted as the vector of probabilities conditional on only the utility indices. However, this general framework allows us to use the same tools to address discrete and continuous choice. For example, choices could involve a single discrete choice, discrete bundle choice, a prospective matching, continuous quantities of several goods, or time use among other settings.

---

[14]The supremum is taken to be $-\infty$ when there is no $\tilde{Y} \in \mathcal{Y}$ such that $\int \tilde{Y}(\varepsilon) d\mu(\varepsilon) = y$.

We places some additional high-level sufficient conditions relative to the conclusions of Lemma 1.

Assumption 2. *Assume the following:*

(i)

$$\overline{Y}(x,\beta) = \underset{y \in \overline{B}}{\operatorname{argmax}} \sum_{k=1}^{K} y_k(\beta_k' x_k) + \overline{D}(y).$$

(ii) $\overline{B} \subseteq \mathbb{R}^K$ *is a nonempty, closed, and convex set.*

(iii) $\overline{D} : \mathbb{R}^K \to \mathbb{R} \cup \{-\infty\}$ *is concave, upper semi-continuous, and finite at some* $y \in \overline{B}$.

Allen and Rehbeck [2019a] provide lower-level conditions that, when combined with Lemma 1, imply this assumption. Part (i) strengthens the conclusion of Lemma 1 to obtain a unique maximizer. Concavity in part (iii) is milder than it first appears, and delivers no additional restrictions on $\overline{Y}(x,\beta)$ when the other assumptions are maintained. See the discussion in Allen and Rehbeck [2019a].

To further present the foundation of the identification results, we present a version of the envelope theorem.

Lemma 2. *Let Assumption 2 hold. It follows that*

$$\overline{Y}_k(x,\beta) = \partial_k V(\beta_1' x_1, \ldots, \beta_K' x_K). \tag{2}$$

Here, $\overline{Y}_k(x,\beta)$ is the k-th component of $\overline{Y}(x,\beta)$ and $\partial_k V(\beta_1' x_1, \ldots, \beta_K' x_K)$ is the derivative with respect to the k-th dimension of V evaluated at the point $(\beta_1' x_1, \ldots, \beta_K' x_K)'$. We use similar notation for the rest of the paper. Differentiability of V is implied by the fact that $\overline{Y}$ is the unique maximizer. This is the primary implication of Assumption 2 that we use for this paper.

Fox et al. [2012] use a structure similar to (2), showing that when V is known, it is possible to identify moments of the distribution of β. We differ because we do not require an analyst to specify V. Instead, we require certain moments to be nonzero as a relevance condition. Appendix B.2 provides further details and a comparison with their approach. A related structure is considered in Lewbel and Pendakur [2017], who identify the distribution of random coefficients when an analogue of $\partial_k V$ is known in

advance or additively separable in arguments. We do not impose this structure.

We also leverage a symmetry property of mixed partial derivatives that results from the optimizing behavior in Assumption 2. For a vector of indices $\gamma = (\gamma_1, \ldots, \gamma_M) \in \{1, \ldots, K\}^M$ and a sufficiently differentiable function $f : \mathbb{R}^K \to \mathbb{R}$, let

$$\partial_\gamma f := \partial_{\gamma_1} \cdots \partial_{\gamma_M} f.$$

Lemma 3. *Suppose V is M-times continuously differentiable in a neighborhood of $\vec{u} \in \mathbb{R}^K$. Let $\gamma, \delta \in \{1, \ldots, K\}^M$ be vectors of indices in which each index occurs the same number of times in both γ and δ. It follows that*

$$\partial_\gamma V(\vec{u}) = \partial_\delta V(\vec{u}).$$

This result states that the order in which we take partial derivatives does not matter. For example, when $M = 2$ we have the usual symmetry property of mixed partial derivatives with respect to dimensions $j, k \in \{1, \ldots, K\}$ that

$$\partial_{j,k} V(\vec{u}) = \partial_{k,j} V(\vec{u}).$$

The lemma follows by repeated application of the $M = 2$ case.

3 Main Result

With the foundations in place, we now turn to the task of identifying moments of random coefficients. We focus on conditions where certain M-th order moments of the distribution of β are identified. In particular, if the assumptions hold for *all M*, then all moments of the distribution of random coefficients are identified.

We assume regressors are continuous and satisfy an exclusion restriction.

Assumption 3. *All covariates are continuous. In addition, each x_k is a vector of regressors specific to the k-th good.*

We now provide some intuition for the main result (Theorem 1). We consider identifying second moments of β ($M = 2$) when there are two goods ($K = 2$) and each

good has a single covariate ($d_k = 1$). We focus on second moments since this example captures the power of the results in the simplest non-trivial setting. We write the partial derivative of a function, f, with respect to the covariates of the j-th good, x_j, as $\partial_{x_j} f$. Differentiating the envelope theorem (Lemma 2) and evaluating at $x = 0$ we obtain

$$\partial_{x_j} \overline{Y}_k(0, \beta) = \partial_{j,k} V(0) \beta_j.^{15}$$

This uses the fact that x_j is continuous and excluded from the utility index of other goods. This can be repeated with other mixed partial derivatives. Importantly, by evaluating derivatives at the point $x = 0$, the terms $\partial_{j,k} V(0)$ do not depend on β. Thus, when integrating over the values of the random coefficients, the term involving V passes outside of the integral. In particular, integrating over β yields the following system of equations

$$
\begin{aligned}
\partial_{x_1} \partial_{x_1} \overline{Y}_2(0) &= \partial_{1,1,2} V(0) \int \beta_1^2 d\nu(\beta) \\
\partial_{x_1} \partial_{x_2} \overline{Y}_2(0) &= \partial_{1,2,2} V(0) \int \beta_2 \beta_1 d\nu(\beta) \\
\partial_{x_2} \partial_{x_1} \overline{Y}_1(0) &= \partial_{2,1,1} V(0) \int \beta_1 \beta_2 d\nu(\beta) \\
\partial_{x_2} \partial_{x_2} \overline{Y}_1(0) &= \partial_{2,2,1} V(0) \int \beta_2^2 d\nu(\beta)
\end{aligned}
\tag{3}
$$

where we have implicitly assumed that differentiation and integration can be interchanged.

Assume that the derivatives of $\overline{Y}$ are identified. At first glance, this is a system of four equations with seven unknowns (clearly the $\beta_1 \beta_2$ and $\beta_2 \beta_1$ moments are equal). However, when V is sufficiently differentiable, partial derivatives of V do not depend on the order of differentiation (Lemma 3), which eliminates two unknowns. Using a scale assumption that $\int \beta_1^2 d\nu(\beta)$ is known *a priori* will eliminate an unknown and gives a system with 4 equations and 4 unknowns. We show that this is enough to identify all second moments of β.

To constructively see how the moments are identified, note that using symmetry of derivatives, the first and third equations identify $\int \beta_1 \beta_2 d\nu(\beta)$. Using this, we

---

[15] Here we abuse notation and for the function f, we let $\partial_s f(0) = \partial_s f(z)|_{z=0}$.

identify $\partial_{1,2,2}V(0)$ using the second equation. Again using symmetry of derivatives and combining this with the last equation identifies $\int \beta_2^2 d\nu(\beta)$. Once all moments are identified, the remaining third order derivatives of V can be identified at 0.[16]

We now provide formal conditions that justify the intuitive argument for any number of goods, covariates, and order of moment M.

Assumption 4. *For the natural number M, $\int \beta_{1,1}^M d\nu(\beta)$ is finite, known a priori, and nonzero.*

Assumption 4 holds if we set $\beta_{1,1} = 1$, for example, but is considerably more general. It allows heterogeneity in the sign of $\beta_{1,1}$, for example. In general, if one wants to identify all moments of β using the main result, then for every M Assumption 4 must hold. This assumption holds when the distribution of $\beta_{1,1}$ is known *a priori* and the distribution has nonzero moments of all orders. If Assumption 4 is dropped, the results in this paper establish identification of the ratio of any nonzero M-th order moments. Thus, Assumption 4 can be appropriately modified by instead holding fixed the value of some other nonzero M-th order moment of the form $\int \beta_{k_1,\ell_1} \cdots \beta_{k_M,\ell_M} d\nu(\beta)$. We show in Section 3.1 that if $\beta_{1,1}$ is independent of all other components of β, then identification is possible using a single scale assumption on the first moment.

Recall that with minor abuse of notation we set

$$\overline{Y}(x) = \int \overline{Y}(x, \beta) d\nu(\beta).$$

We require the following regularity conditions.

Assumption 5. *For the natural number M, the following conditions hold:*

(i) For each good k, one can interchange integration and differentiation for all M-th

---

[16]This part also requires the equations

$$\partial_{x_1} \partial_{x_1} \overline{Y}_1(0) = \partial_{1,1,1} V(0) \int \beta_1^2 d\nu(\beta)$$

$$\partial_{x_2} \partial_{x_2} \overline{Y}_2(0) = \partial_{2,2,2} V(0) \int \beta_2^2 d\nu(\beta)$$

to identify $\partial_{1,1,1}V(0)$ and $\partial_{2,2,2}V(0)$.

order partial derivatives at $x = 0$ so that

$$\partial_{x_{k_1},\ell_1} \cdots \partial_{x_{k_M},\ell_M} \overline{Y}_k(0) = \int \partial_{x_{k_1},\ell_1} \cdots \partial_{x_{k_M},\ell_M} \overline{Y}_k(0, \beta) d\nu(\beta)$$

holds.

(ii) Each M-th order moment

$$\int \beta_{k_1,\ell_1} \cdots \beta_{k_M,\ell_M} d\nu(\beta)$$

exists and is finite.

(iii) V is $(M + 1)$-times continuously differentiable in a neighborhood of 0.

(iv) For each $\gamma \in \{1, \ldots, K\}^{M+1}$,

$$\partial_\gamma V(0) \neq 0.$$

(v) $\overline{Y}(x)$ is known in a neighborhood of $x = 0$, or more generally it is known in a neighborhood of $x = 0$ with respect to the weakly positive orthant of $\mathbb{R}^{\sum_{k=1}^K d_k}$.[17]

These regularity conditions parallel assumptions in Fox et al. [2012]. To interpret part (i), note that ν can be a discrete probability measure over β with finite support. For discrete measures, (i) holds whenever $\overline{Y}_k(x, \beta)$ is M-times differentiable in x for every β in its support. Part (ii) formalizes that the moments we wish to identify exist and are finite.

Parts (iii) and (iv) can be linked to derivatives of the function $\overline{Y}(x)$ via the envelope theorem (Lemma 2). Indeed, differentiating the envelope theorem for the k-th good, evaluating the derivative of $\bar{Y}_k$ with respect to $x_{j,\ell}$ at $x = 0$, and taking expectations yields

$$\frac{\partial \overline{Y}_k(0)}{\partial x_{j,\ell}} = \partial_{j,k} V(0) \int \beta_{j,\ell} d\nu(\beta). \tag{4}$$

Thus, when V is $(M + 1)$-times continuously differentiable, it follows that $\overline{Y}$ is M-times continuously differentiable. Moreover, if one sees empirically that $\frac{\partial \overline{Y}_k(0)}{\partial x_{j,\ell}} \neq 0$,

---

[17]More formally, the second condition can be written as follows: for some neighborhood H of $x = 0$ in the usual topology on $\mathbb{R}^{\sum_{k=1}^K d_k}$, $\overline{Y}(x)$ is known on $H \cap \mathbb{R}_+^{\sum_{k=1}^K d_k}$, where $\mathbb{R}_+ = \mathbb{R} \cap [0, \infty)$.

then it follows that that $\partial_{j,k}V(0) \neq 0$ (whenever this derivative exists).

Fox et al. [2012] show that condition (iv) holds for random coefficients logit, for "most" values of nonrandom intercepts. Specifically, the set of intercepts that violate (iv) for some γ has Lebesgue measure 0. In general, whether (iv) holds depends on features of the distribution of ε and choice of D, which are example specific. For example, part (iv) rules out pure characteristic discrete choice models as in Berry and Pakes [2007] and Dunker et al. [2017]. These models do not include a random intercept, and so the value function for the pure characteristics model, V^{PC}, can be written as

$$V^{PC}(\beta'_1 x_1, \ldots, \beta'_K x_K) = \sup_{y \in \overline{B}} \sum_{k=1}^{K} y_k(\beta'_k x_k)$$

without the additive disturbance $\overline{D}$, where $\overline{B}$ is the probability simplex. V^{PC} does not have a non-zero derivative at $x = 0$ for any $M \geqslant 1$ with this constraint set. This choice of V^{PC} also does not always induce a unique maximizer.

More generally, condition (iv) requires that goods in the demand system are related. For example, if the original K-good demand system can be written as K separate 1-good demand systems, then derivatives of the form $V_{j,k}(0)$ will be zero for $j \neq k$. This is because under this separability assumption, the utility index of good j does not alter the demand for the k-th good. In general, (iv) cannot be relaxed for the main result to hold without additional assumptions. For example, if we impose that $\beta_j = \beta_k$ (a.s.) for all $j, k \in \{1, \ldots, K\}$, then one can identify ratios of moments under the weaker assumption that $\partial_j^{M+1} V(0) \neq 0$ for some j. See Appendix B.3.

Condition (v) states that $\overline{Y}(x)$ is identified over a small region near $x = 0$. The constructive identification results in Fox et al. [2012] and Chernozhukov et al. [2019a] have also made use of variation around zero. In contrast, most of the literature instead requires identification of $\overline{Y}$ either for all x or for a set over which x is unbounded along some dimensions.

To interpret condition (v), suppose that X, β, and ε are all independent, and we identify $\overline{Y}$ from a continuous version of the conditional mean of Y given X. For this case, condition (v) is implied when the support of X contains an open ball around $x = 0$. The second more general part of (v) highlights that the results also apply when the average structural function is identified over a weakly positive region. Thus,

our results do not rule out prices. We can handle this case because we only need to identify certain derivatives of $\overline{Y}$ at 0. These derivatives of $\overline{Y}$ at 0 are identified in this case by calculating derivatives from "one-sided" limits involving non-negative numbers.

The final assumption used for identification is that a sufficiently rich set of M-th order moments of β are nonzero.

Assumption 6. *For the natural number M and each tuple of good indices $(k_1, \ldots, k_M) \in \{1, \ldots, K\}^M$, there is a corresponding tuple of characteristic indices $(\ell_1, \ldots, \ell_M) \in \prod_{m=1}^{M} \{1, \ldots, d_{k_m}\}$ such that the M-th order moment*

$$\int \beta_{k_1, \ell_1} \cdots \beta_{k_M, \ell_M} \nu(d\beta)$$

exists and is nonzero.

This is a relevance condition. It is not necessary to know which indices $(\ell_1, \ldots, \ell_M)$ satisfy this condition in advance.[18] A sufficient condition for this is that for every k-th good there is a regressor $\ell_k \in \{1, \ldots, d_k\}$ such that either $\beta_{k,\ell_k} \geqslant 0$ almost surely or $\beta_{k,\ell_k} \leqslant 0$ almost surely, with positive probability that the inequality is strict. A stronger condition that implies this is that $\beta_{k,1} = 1$ (a.s.) for every k-th good by setting $\ell_1 = \cdots = \ell_M = 1$. This is a common assumption in the literature [Berry and Haile, 2009, Briesch et al., 2010, Dunker et al., 2017]. However, Ichimura and Thompson [1998] and Gautier and Kitamura [2013] establish identification of random coefficients models for binary discrete choice using a more general halfspace condition.

With these assumptions, we can now state the main result of the paper.

Theorem 1. *Let Assumptions 1-6 hold with the same natural number M. Each M-th order moment of the form*

$$\int \beta_{k_1, \ell_1} \cdots \beta_{k_M, \ell_M} d\nu(\beta)$$

is identified. In addition, for each $\gamma \in \{1, \ldots, K\}^{M+1}$,

$$\partial_\gamma V(0)$$

is identified.

---

[18]See the discussion after Lemma A.2 in Appendix A.2.

This result establishes nonparametric identification of certain moments of β. It can be directly used to establish semiparametric identification of the distribution of β for certain parametric families without specifying other objects (e.g. V). For example, if β is normally distributed then Theorem 1 identifies the distribution when the assumptions hold for $M \in \{1, 2\}$ because normal distributions are characterized by means and covariances. Recall that while we identify non-centered moments, we can use this information to identify centered moments. More generally, for any distribution of β that is defined by its moments up to order M this result estabilishes identification of the distribution.

Corollary 1. *Suppose Assumptions 1-6 hold for each $M \leqslant \overline{M}$ (which could be ∞) and the distribution of β is determined by its first $\overline{M}$ moments. It follows that the distribution of β is identified.*

Fox et al. [2012] and il Kim [2014] describe a sufficient condition for a distribution to be determined by its moments. Distributions with compact finite support are determined by their moments. Lognormal distributions are an example of distributions that are not determined by integer moments [Heyde, 1963]. That is, there are other nonparametric distributions that can match the same moments. However, the parameters may still be identified within the lognormal class.

Remark 1 (Constructive Identification). The proof of Theorem 1 is constructive. While a detailed approach to estimation is beyond the scope of this paper, the constructive results can be used to show how consistent estimation of the ratios of certain derivatives allows one to consistently estimate moments of the distribution of random coefficients. See Appendix B.1 for a brief outline. Estimation error of the M-th order moments can be bounded when the analyst has a suitable estimator of the M-th order derivatives of the average structural function. For example, Chen and Christensen [2018] provide an approach to estimate derivatives of average structural functions in the presence of endogeneity.

3.1 Independence of $\beta_{1,1}$

Identifying the distribution of β using Corollary 1 requires Assumption 4, which specifies all moments of the form $\int \beta_{1,1}^M d\nu(\beta)$. When the distribution of β is identified from its moments, one must specify the marginal distribution of $\beta_{1,1}$ in advance to apply

Corollary 1. While the common assumption $\beta_{1,1} = 1$ (a.s.) implies Assumption 4, one may not want to impose either this assumption or the weaker assumption that $\beta_{1,1}$ has a known distribution. This section describes an alternative assumption that ensures identification of moments of β. In particular, we assume $\beta_{1,1}$ is independent of other components of β.

With this independence assumption, we show that a single scale assumption on the first moment of $\beta_{1,1}$ allows us to identify a rich collection of moments. This contrasts with Theorem 1, which uses an assumption on the M-th order moment of $\beta_{1,1}$ to identify *only* M-th order moments of β.

Assumption 7. *For the measure ν as defined in Assumption 1, $\beta_{1,1}$ is independent of all other components of β. In addition, $|\int \beta_{1,1} d\nu(\beta)|$ is finite, known a priori, and nonzero.*

Alternatively, one could set the absolute value of some other order moment of $\beta_{1,1}$, but we focus on the first moment since it facilitates interpretation. Independence between $\beta_{1,1}$ and other components is considerably weaker than assuming $\beta_{1,1} = 1$ almost surely. For example, this allows $\beta_{1,1}$ to be sometimes negative and sometimes positive. Thus, different individuals can be repelled or attracted to higher values of $x_{1,1}$.

Replacing Assumption 4 with Assumption 7, we obtain the following counterpart of Theorem 1.

Proposition 1. *Let $K \geqslant 2$ and Assumptions 1-3 and 5-7 hold for all for each $M \leqslant \overline{M}$ (which could be ∞). It follows that the M-th order moment of the form*

$$\int \beta_{k_1,\ell_1} \cdots \beta_{k_M,\ell_M} \nu(d\beta)$$

is identified. In addition, for each $\gamma \in \{1,\ldots,K\}^{M+1}$,

$$\partial_\gamma V(0)$$

is identified.

If $K = 1$, the same conclusions hold given the additional assumption that for each

each $M \leqslant \overline{M}$, there exists an order $M - 1$ moment such that

$$\int \beta_{1,\ell_1} \cdots \beta_{1,\ell_{M-1}} d\nu(\beta) \neq 0,$$

where $\ell_m \neq 1$ for every $m \in \{1, \ldots, M - 1\}$.

Relative to Theorem 1, independence of $\beta_{1,1}$ from the other components allows us to relate the M-th and $M - 1$ order moments. To see this, consider some M-th order moment in which $\beta_{1,1}$ appears exactly once. Using independence, we obtain

$$\int \beta_{1,1} \beta_{k_2,\ell_2} \cdots \beta_{k_M,\ell_M} d\nu(\beta) = \int \beta_{1,1} d\nu(\beta) \int \beta_{k_2,\ell_2} \cdots \beta_{k_M,\ell_M} d\nu(\beta).$$

In the proof, we show that $\int \beta_{1,1} d\nu(\beta)$ can be identified when $|\int \beta_{1,1} d\nu(\beta)|$ is finite, known *a priori*, and non-zero. With this knowledge, we can identify the ratio of all M-th order moments to all $(M-1)$-th order moments and apply induction to identify all $M \leqslant \overline{M}$ order moments.

Remark 2 (One good). When $K \geqslant 2$, the assumptions of Proposition 1 impose that there are multiple relevant characteristics. The additional assumption in the $K = 1$ case is a relevance condition on a characteristic other than $x_{1,1}$.

Remark 3 ("Normalizations"). Provided $\int \beta_{1,1} \nu(d\beta)$ exists and is nonzero, it is a normalization to set $|\int \beta_{1,1} \nu(d\beta)| = 1$. In other words, this imposes no additional restrictions on the model. This is seen by noting that if we divide the original latent utility model by $|\int \beta_{1,1} \nu(d\beta)|$, then the argmax set does not change and none of the assumptions in Proposition 1 are affected by this division.

A natural intuition is that when $\beta_{1,1} > 0$ almost surely, it is also a normalization to divide the latent utility by $\beta_{1,1}$ and rewrite the problem with $\beta_{1,1} = 1$. This is true if we only inspect the original latent utility model (Equation 1), but is no longer true when we consider independence or certain other additional assumptions on the integrated model in Assumption 2. Recall that Assumption 1 means β and ε are independent in the definition of the average structural function $\overline{Y}$. In general, this assumption is not invariant to division by $\beta_{1,1}$. This means that setting $\beta_{1,1} = 1$ (a.s.) provides additional restrictions relative to the assumptions of Proposition 1.

4   Welfare Analysis and Counterfactuals

We now turn to identification of certain welfare and counterfactual objects. Identification is established given identification of certain features of V, which is the indirect utility function obtained when random intercepts are integrated out. We first provide three results that identify differences in V. Using these results, we discuss welfare analysis and counterfactuals.

The reason we identify V is that we can use the envelope theorem to determine certain average choices

$$\overline{Y}(x, \beta) = \nabla V(\beta_1' x_1, \ldots, \beta_K' x_K).$$

We require identification of the right hand side at values other than 0 to consider counterfactuals at new values of covariates.

4.1   Identification of V

We first provide conditions under which identification of partial derivatives of V at 0 allows us to directly extrapolate the function. Specifically, we assume V is a real analytic function. That is, V has derivatives of all orders and agrees with its Taylor series in a neighborhood of every point. Real analytic functions have the important property that local information can be used to reconstruct the function globally by extrapolating. This is similar to common parametric classes of functions. However, the set of real analytic functions is infinite dimensional.

Corollary 2. *Let the assumptions of Theorem 1 or the assumptions of Proposition 1 hold with $\overline{M} = \infty$. If V is a real analytic function, then it is identified up to an additive constant.*

One way to drop the assumption that V is a real analytic function is to instead assume $\beta_{k,1} = 1$ almost surely for each k. With this assumption, let $\tilde{x}$ be a value that is zero for every characteristic except the first characteristic of each good. Then the envelope theorem (Lemma 2) specializes to

$$\overline{Y}_k(\tilde{x}, \beta) = \partial_k V(\tilde{x}_{1,1}, \ldots, \tilde{x}_{K,1}).$$

This does not depend on β, and so by taking expectations, the average structural function identifies the derivative of V at the point $(\tilde{x}_{1,1}, \ldots, \tilde{x}_{K,1})$. By integrating the derivatives we can identify differences in V, as we now formalize.

Proposition 2. *Let Assumption 2 hold and assume $\beta_{k,1} = 1$ almost surely for each $k \in \{1, \ldots, K\}$. Suppose $\overline{Y}(x)$ is identified for all $x = (x'_1, \ldots, x'_K)'$ satisfying $x_{k,1} \in [\underline{x}_{k,1}, \overline{x}_{k,1}]$ for each k, and $x_{k,j} = 0$ for $j > 1$. Then differences in V are identified over the region $\times_{k=1}^{K} [\underline{x}_{k,1}, \overline{x}_{k,1}]$. In particular, if $\underline{x}_{k,1} = -\infty$ and $\overline{x}_{k,1} = \infty$ for each k, then V is identified up to an additive constant.*

The results on welfare and counterfactual analysis require that derivatives of V be identified at certain values $(\beta'_1 x_1, \cdots, \beta'_K x_K)$. If the support of β is compact, then it is not necessary to identify V everywhere, and so it is not necessary to have $\underline{x}_{k,1} = -\infty$ and $\overline{x}_{k,1} = \infty$ to apply Proposition 2.

Finally, we mention a third way to identify differences in V. A key distinction is that it requires identification of the distribution of $\overline{Y}(x, \beta)$ for fixed x, rather than identification of the average structural function as in the rest of the paper. We adapt the following lemma.

Lemma 4 (McCann [1995]; statement from Chernozhukov et al. [2019b], Corollary 2)**.** *Let $W = f(\eta)$, where W and η have the same finite dimension. Suppose f is the gradient of a convex function, η has a known distribution that is absolutely continuous with respect to Lesbesgue measure, the distribution of W is known, and η and W have finite variance. It follows that f is identified.*

This result can be applied to our setting by adapting the envelope theorem,

$$\overline{Y}(x, \beta) = \nabla V(\beta'_1 x_1, \ldots, \beta'_K x_K).$$

Interpret $W = \overline{Y}(x, \beta)$, $f = \nabla V$, and $\eta = (\beta'_1 x_1, \ldots, \beta'_K x_K)$. When x is fixed and the distribution of β is identified (from previous arguments), the distribution of η is known. The function V is convex, and so the lemma provides conditions under which ∇V is identified. Importantly, the lemma can be applied at a *single* x, so it is not necessary to have full support of covariates to apply the result. Such an x cannot be arbitrary. For example, when $x = 0$ the distribution corresponding to η is not absolutely continuous. The lemma can still be applied for x near but not equal to 0. Moreover, to apply the lemma, the distribution of β cannot be degenerate, i.e. there

must be truly "random" coefficients. If β is almost surely equal to a constant, then the distribution corresponding to η is not absolutely continuous and the lemma does not apply.

Importantly, to apply Lemma 4 in our setting, the distribution of $\overline{Y}(x, \beta)$ must be identified at some fixed x. One example in which this lemma can be applied is when ε in the original latent utility model is not present, so that $\overline{Y}(x, \beta)$ corresponds to the observable choices given x and β. Such structure could be appropriate in a continuous choice model in which all unobservable heterogeneity is controlled by the random slopes β, and in which the choices (rather than e.g. average choices for a group of individuals) are observed.

4.2   Welfare

We now describe how identification of V leads to identification of certain welfare objects. First, recall that V may be interpreted as the indirect utility conditional on the utility index (Lemma 1) where the random intercept ε under the measure μ is integrated out. To interpret $V(\cdot)$ as a welfare object, suppose β is an individual-specific term that is random across the population but constant across decisions for the same individual. Interpret the random intercept ε as an idiosyncratic taste shock across decision problems. Then $V(\beta_1' x_1, \ldots, \beta_K' x_K)$ is an individual-specific (integrated) indirect utility. The conditions of Corollary 1 identify the distribution of $V(\beta_1' x_1, \ldots, \beta_K' x_K)$ under the measure ν up to an additive constant once the values of covariates are fixed.

Thus, we can identify the distribution of individual-specific indirect utilities. By further integrating out the distribution of β, we also identify differences in the average indirect utility via Lemma 1:

$$\int V(\beta_1' x_1, \ldots, \beta_K' x_K) d\nu(\beta) = \int \int \sup_{y \in B} \sum_{k=1}^{K} y_k(\beta_k' x_k) + D(y, \varepsilon) d\mu(\varepsilon) d\nu(\beta). \quad (5)$$

This result holds regardless of whether the distribution of ε is identified since V is a welfare-relevant summary measure of the distribution of ε. Indeed, we do not

establish identification of the distribution of

$$\sup_{y \in B} \sum_{k=1}^{K} y_k(\beta_k' x_k) + D(y, \varepsilon)$$

according to the product measure $\mu \times \nu$ over (β, ε). In particular, since this paper does not study identification of ε, we do not identify the distribution of indirect utilities including random intercepts.

Note that the units of (5) are relative to the scale assumption used to identify the distribution of β. If we impose the scale assumptions in Theorem 1 to apply Corollary 2, then the distribution of $\beta_{1,1}$ is fixed. Thus, the units of Equation 5 are set by the *distribution* of the conversion rate between $x_{1,1}$ and utils. In contrast, if we impose the conditions of Proposition 1, then the scale is determined by $\left| \int \beta_{1,1} \nu(\beta) \right|$. For this case, the units of Equation 5 are relative to the *average* conversion rate between $x_{1,1}$ and utils.

An alternative measure of average indirect utility is

$$\int \frac{1}{|\beta_{1,1}|} V(\beta_1' x_1, \dots, \beta_K' x_K) d\nu(\beta) = \int \int \frac{1}{|\beta_{1,1}|} \sup_{y \in B} \sum_{k=1}^{K} y_k(\beta_k' x_k) + D(y, \varepsilon) d\mu(\varepsilon) d\nu(\beta).$$

$$(6)$$

This sets the conversion rate of $x_{1,1}$ and utils to ± 1. Importantly, this preserves whether the first characteristic is desirable or undesirable. It also forces the intensity of preference to be constant across individuals. This welfare measure is most interpretable when the regressor has a homogeneous sign. For example, if $x_{1,1}$ is the (negative) price of good 1, then $\beta_{1,1} < 0$ is a natural assumption, and the units of Equation 6 are in dollars.

4.3   Counterfactuals

Once V is identified, we can also answer certain counterfactual questions involving quantities at new values of covariates. To this end, recall from Lemma 2 that

$$\overline{Y}_k(x, \beta) = \partial_k V(\beta_1' x_1, \dots, \beta_K' x_K).$$

$$(7)$$

24

Here, $\overline{Y}_k(x, \beta)$ is the demand for good k fixing covariates and the random intercept, but integrating out the distribution of ε. We interpret β as an individual-specific parameter that is constant across decision problems, while ε is an idiosyncratic shock that can vary across decision problems. Thus, $\overline{Y}_k(x, \beta)$ is the individual-specific average quantity of the k-th good. Once V and the distribution of β are identified, we can identify the distribution of $\overline{Y}_k(x, \cdot)$ from Equation 7.

Conceptually, this shows it is possible to start with identification of the average structural function ("mean choices") around $x = 0$ to identify the integrated choices $\overline{Y}_k(x, \beta)$ at all values of the covariates. This also implies that any value x at which $\overline{Y}$ can be identified directly from data (as opposed to the theoretical analysis just described) provides overidentifying information.

5 Discussion

We now provide additional discussion of the main results in the paper.

Remark 4 (Location of Taste Homogeneity). Our analysis can be applied to settings in which covariates do not vary around zero. For example, we could recenter via $\tilde{X} = X - \mathbb{E}[X]$ so that identification uses variation in the average structural function around the mean, rather than variation around 0. This is noted as well in Fox et al. [2012]. Importantly, the assumptions in this paper are not typically invariant to recentering. Thus, the assumptions must be made *given* a particular centering at which the average structural function is identified. In addition, the location of recentering defines where the random slopes do not alter preferences since $\beta'x = 0$ when $x = 0$. In words, the choice of centering sets a location of taste homogeneity with regard to the random slopes (but not random intercepts).

Remark 5 (Nonlinear Random Coefficients). The results may be adapted to certain models in which coefficients are not linear. Suppose instead of the linear index $\beta'_k x_k$, we have $x_k^{\rho_k}$ for a scalar shifter x_k and random exponent ρ_k. Applying the envelope theorem yields

$$\partial_{x_j} \overline{Y}_k(x, \rho) = \partial_{j,k} V(x_1^{\rho_1}, \ldots, x_K^{\rho_K}) \rho_j x_j^{\rho_j - 1}$$

where $\rho = (\rho_1, \ldots, \rho_K)$ collects the random exponents. In particular, the partial derivatives of V can be evaluated at a vector of ones so that $x_j^{\rho_j - 1} = 1$ for any random exponent. By taking expectations with respect to ρ, evaluating the above equation around covariates equal to one, and using symmetry of mixed partial derivatives, we obtain

$$\frac{\partial_{x_j} \overline{Y}_k(1)}{\partial_{x_k} \overline{Y}_j(1)} = \frac{\mathbb{E}[\rho_j]}{\mathbb{E}[\rho_k]}.$$

Ratios of higher order moments can be identified by considering additional derivatives of the average structural function, and by using symmetry of mixed partials of V evaluated at the vector of ones.

Remark 6 (Complementarity and Derivatives of V). Recall that the envelope theorem yields

$$\partial_{x_j} \overline{Y}_k(0, \beta) = \partial_{j,k} V(0) \beta_j.$$

Thus, second order mixed partial derivatives of V describe how changes in the utility index of good x_j alter the demand for good k. The sign of $\partial_{j,k} V(0)$ describes whether goods are local complements ($\partial_{j,k} V(0) \geqslant 0$) or substitutes ($\partial_{j,k} \leqslant V(0)$). Theorem 1 provides conditions under which this derivative is identified, and thus we obtain information on complementarity/substitability with random coefficients. Several papers have studied complementarity in the bundles model (Example 2) when characteristics shift the marginal utility of a good homogenously (Gentzkow [2007], Fox and Lazzati [2017], Chernozhukov et al. [2015], Allen and Rehbeck [2019d,c]). To our knowledge, the only paper that studies identification of complementarity in the bundles model with heterogeous tastes for characteristics is Dunker et al. [2017]. Rather than working with only the average structural function ("mean demands"), Dunker et al. [2017] require identification of the frequencies that goods are chosen together.

Remark 7 (Homogeneity of Coefficients). We do not require the assumption that coefficients are the same across goods, $\beta_j = \beta_k$. One interpretation of this assumption is that preferences are driven by observable characteristics [Gorman, 1980, Lancaster, 1966], not the label of the good (j vs k). In this paper, we assume that the shifters associated with β_k vary for good k. Thus, setting $\beta_j = \beta_k$ means the shifters that vary for good j are the same as for good k. This assumption is inconsistent with

some empirical settings, especially outside of discrete choice. For example, it is not satisfied for the bundles model of Gentzkow [2007] where internet speed varies for online news but not print news.

Placing restrictions relating coefficients across different goods allows one to either weaken the conditions used for identification or use alternative techniques. See Appendix B.3 for additional discussion and relation to Chernozhukov et al. [2019a].

Remark 8 (Testability). The conditions of Theorem 1 imply testable implications because of theoretical relationships between different moments. To see this, we revisit the system of equations (3) used previously to illustrate the identification technique. Dividing the first and third equations, and the second and fourth equations, we obtain

$$\frac{\partial_{x_1}\partial_{x_1}\overline{Y}_2(0)}{\partial_{x_2}\partial_{x_1}\overline{Y}_1(0)} = \frac{\int \beta_1^2 \nu(d\beta)}{\int \beta_1\beta_2 \nu(d\beta)} \qquad\qquad \frac{\partial_{x_2}\partial_{x_2}\overline{Y}_1(0)}{\partial_{x_1}\partial_{x_2}\overline{Y}_2(0)} = \frac{\int \beta_2^2 \nu(d\beta)}{\int \beta_2\beta_1 \nu(d\beta)}.$$

Multiplying these equations and using the Cauchy-Schwarz inequality yields the testable restriction

$$\frac{\partial_{x_1}\partial_{x_1}\overline{Y}_2(0)}{\partial_{x_2}\partial_{x_1}\overline{Y}_1(0)} \cdot \frac{\partial_{x_2}\partial_{x_2}\overline{Y}_1(0)}{\partial_{x_1}\partial_{x_2}\overline{Y}_2(0)} \geq 1.$$

Note that this inequality only concerns the average structural function close to 0.

Remark 9 (Determinacy by Moments). The results in this paper establish identification of certain moments of the distribution of random coefficients. It is natural to wonder whether the distribution can be identified without requiring that it be uniquely determined by its moments. A related question has been studied previously. Specifically, consider the setting

$$Y = A + B'Z,$$

where Z is independent of (A, B) and the distribution of (Y, Z) is identified. Building on Belisle et al. [1997], Masten [2017] shows that, if the support of Z is bounded, then identification of either (A, B) or even the marginal distributions B_k requires that the distribution (A, B) is determined by its moments. In light of this, it is possible that given only the conditions of Theorem 1 for each M, identification of the distribution of β requires that it be determined by its moments. Recall that we only assume identification of the average structural function $\overline{Y}(x)$ in a bounded set around 0.

References

Victor H Aguiar. Random categorization and bounded rationality. *Economics Letters*, 159:46–52, 2017.

Roy Allen. Injectivity and the law of demand. *Available at SSRN 3437946*, 2019.

Roy Allen and John Rehbeck. Identification with additively separable heterogeneity. *Econometrica*, 87(3):1021–1054, 2019a.

Roy Allen and John Rehbeck. Revealed stochastic choice with attributes. *Available at SSRN 2818041*, 2019b.

Roy Allen and John Rehbeck. Latent complementarity in bundles models. *Working Paper*, 2019c.

Roy Allen and John Rehbeck. Hicksian complementarity and perturbed utility models. *Working Paper*, 2019d.

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

Simon P Anderson, André De Palma, and J-F Thisse. A representative consumer theory of the logit model. *International Economic Review*, pages 461–466, 1988.

Simon P Anderson, Andre De Palma, and Jacques François Thisse. *Discrete choice theory of product differentiation*. Cambridge, MA: MIT press, 1992.

Claude Belisle, Jean-Claude Massé, Thomas Ransford, et al. When is a probability measure determined by infinitely many projections? *The Annals of Probability*, 25 (2):767–786, 1997.

Steven Berry and Ariel Pakes. The pure characteristics demand model. *International Economic Review*, 48(4):1193–1225, 2007.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, pages 841–890, 1995.

Steven T Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

Steven T Berry and Philip A Haile. Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Technical report, National Bureau of Economic Research, 2009.

Steven T Berry and Philip A Haile. Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797, 2014.

Richard Blundell and James L Powell. Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs*, 36:312–357, 2003.

J Hayden Boyd and Robert E Mellman. The effect of fuel economy standards on the us automotive market: an hedonic demand analysis. *Transportation Research Part A: General*, 14(5-6):367–378, 1980.

Richard L Brady and John Rehbeck. Menu-dependent stochastic feasibility. *Econometrica*, 84(3):1203–1223, 2016.

Richard A Briesch, Pradeep K Chintagunta, and Rosa L Matzkin. Nonparametric discrete choice models with unobserved heterogeneity. *Journal of Business & Economic Statistics*, 28(2):291–307, 2010.

N Scott Cardell and Frederick C Dunbar. Measuring the societal impacts of automobile downsizing. *Transportation Research Part A: General*, 14(5-6):423–434, 1980.

Xiaohong Chen and Timothy M Christensen. Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric iv regression. *Quantitative Economics*, 9(1):39–84, 2018.

Victor Chernozhukov, Whitney K Newey, and Andres Santos. Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311*, 2015.

Victor Chernozhukov, Iván Fernández-Val, and Whitney K Newey. Nonseparable multinomial choice models in cross-section and panel data. *Journal of Econometrics*, 211(1):104–116, 2019a.

Victor Chernozhukov, Alfred Galichon, Marc Henry, and Brendan Pass. Single market nonparametric identification of multi-attribute hedonic equilibrium models. *arXiv preprint arXiv:1709.09570*, 2019b.

Khai Chiong, Yu-Wei Hsieh, and Matthew Shum. Counterfactual estimation in semiparametric discrete-choice models. Technical report, 2017. URL http://dx.doi.org/10.2139/ssrn.2979446.

Khai Xiang Chiong, Alfred Galichon, and Matt Shum. Duality in dynamic discrete-choice models. *Quantitative Economics*, 7(1):83–115, 2016.

Christopher T Conlon and Julie Holland Mortimer. Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4): 1–30, 2013.

Angus Deaton and John Muellbauer. An almost ideal demand system. *The American Economic Review*, 70(3):312–326, 1980.

Xavier D'Haultfœuille and Philippe Février. Identification of nonseparable triangular models with discrete instruments. *Econometrica*, 83(3):1199–1210, 2015.

T Domenich and D McFadden. Urban travel demand: a behavioural approach. *Worth-Holland, Amsterdam*, 1975.

Fabian Dunker, Stefan Hoderlein, and Hiroaki Kaido. Nonparametric identification of endogenous and heterogeneous aggregate demand models: complements, bundles and the market level. *Working Paper*, 2017.

Fabian Dunker, Stefan Hoderlein, Hiroaki Kaido, and Robert Sherman. Nonparametric identification of the distribution of random coefficients in binary response static games of complete information. *Journal of Econometrics*, 206(1):83–102, 2018.

Ivar Ekeland, James J Heckman, and Lars Nesheim. Identifying hedonic models. *American Economic Review*, 92(2):304–309, 2002.

Ivar Ekeland, James J Heckman, and Lars Nesheim. Identification and estimation of hedonic models. *Journal of political economy*, 112(S1):S60–S109, 2004.

Kfir Eliaz and Ran Spiegler. Consideration sets and competitive marketing. *The Review of Economic Studies*, 78(1):235–262, 2011.

Daniel Ershov, Jean-William Laliberté, and Scott Orr. Mergers in a model with complementarity. Technical report, 2018. Working Paper.

Mogens Fosgerau, Julien Monardo, and André De Palma. The inverse product differentiation logit model. 2019.

Jeremy T Fox. A note on nonparametric identification of distributions of random coefficients in multinomial choice models. Technical report, National Bureau of Economic Research, 2017.

Jeremy T Fox and Amit Gandhi. Nonparametric identification and estimation of random coefficients in multinomial choice models. *The RAND Journal of Economics*, 47(1):118–139, 2016.

Jeremy T Fox and Natalia Lazzati. A note on identification of discrete choice models for bundles and binary games. *Quantitative Economics*, 8(3):1021–1036, 2017.

Jeremy T Fox, Kyoo il Kim, Stephen P Ryan, and Patrick Bajari. The random coefficients logit model is identified. *Journal of Econometrics*, 166(2):204–212, 2012.

Drew Fudenberg, Ryota Iijima, and Tomasz Strzalecki. Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409, 2015.

Xavier Gabaix. Chapter 4 - behavioral inattention. In B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, editors, *Handbook of Behavioral Economics - Foundations and Applications 2*, volume 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, pages 261 – 343. North-Holland, 2019. doi: https://doi.org/10.1016/bs.hesbe.2018.11.001. URL http://www.sciencedirect.com/science/article/pii/S2352239918300216.

Alfred Galichon. *Optimal transport methods in economics*. Princeton University Press, 2018.

Alfred Galichon and Bernard Salanié. Cupid's invisible hand: Social surplus and

identification in matching models. Technical report, 2015. URL `http://dx.doi.org/10.2139/ssrn.2979446`.

Eric Gautier and Yuichi Kitamura. Nonparametric estimation in random coefficients binary choice models. *Econometrica*, 81(2):581–607, 2013.

Matthew Gentzkow. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, 97(3):713–744, 2007.

William M Gorman. A possible procedure for analysing quality differentials in the egg market. *The Review of Economic Studies*, 47(5):843–856, 1980.

Florian Gunsilius. Nonparametric point-identification of multivariate models with binary instruments. *Working Paper*, 2019.

Jerry A Hausman and David A Wise. A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica: Journal of the Econometric Society*, pages 403–426, 1978.

James J Heckman. Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of political Economy*, 109(4):673–748, 2001.

James J Heckman, Rosa L Matzkin, and Lars Nesheim. Nonparametric identification and estimation of nonadditive hedonic models. *Econometrica*, 78(5):1569–1591, 2010.

Chris C Heyde. On a property of the lognormal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 25(2):392–393, 1963.

Josef Hofbauer and William H Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.

Alessandro Iaria and Ao Wang. Identification and estimation of demand for bundles. Technical report, 2019. Working Paper.

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

Kyoo il Kim. Identification of the distribution of random coefficients in static and dy-namicdiscrete choice models. *The Korean Economic Review*, 30(2):191–216, 2014.

Guido W Imbens and Whitney K Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512, 2009.

Nail Kashaev. Identification of semiparametric discrete outcome models with bounded covariates. *arXiv preprint arXiv:1811.05555*, 2018.

Kelvin J Lancaster. A new approach to consumer theory. *Journal of political economy*, 74(2):132–157, 1966.

Arthur Lewbel and Krishna Pendakur. Unobserved preference heterogeneity in de-mand using generalized random coefficients. *Journal of Political Economy*, 125(4):1100–1148, 2017.

Charles F Manski. The structure of random utility models. *Theory and decision*, 8(3):229–254, 1977.

Paola Manzini and Marco Mariotti. Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176, 2014.

Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y Ozbay. Revealed attention. *American Economic Review*, 102(5):2183–2205, 2012.

Matthew A Masten. Random coefficients on endogenous variables in simultaneous equations models. *The Review of Economic Studies*, 85(2):1193–1250, 2017.

Rosa L Matzkin. Restrictions of economic theory in nonparametric methods. *Handbook of Econometrics*, 4:2523–2558, 1994.

Rosa L Matzkin. Constructive identification in some nonseparable discrete choice models. *Journal of Econometrics*, 211(1):83–103, 2019.

Robert J McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–324, 1995.

Daniel McFadden. Econometric models of probabilistic choice. In Charles F. Manski

and Daniel McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 198–272. Cambridge, MA: MIT Press, 1981.

Daniel L McFadden and Mogens Fosgerau. A theory of the perturbed consumer with general budgets. Technical report, National Bureau of Economic Research, 2012.

Julien Monardo. The flexible inverse logit (fil) model. *Available at SSRN 3388972*, 2019.

Aviv Nevo. A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of economics & management strategy*, 9(4):513–548, 2000.

Whitney K Newey and James L Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

R Tyrrell Rockafellar. *Convex analysis*. Number 28. Princeton university press, 1970.

Xiaoxia Shi, Matthew Shum, and Wei Song. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Econometrica*, 86(2):737–761, 2018.

Alexander Torgovitsky. Identification of nonseparable models using instruments with small support. *Econometrica*, 83(3):1185–1197, 2015.

Kenneth E Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2009.

Ao Wang. A blp demand model of product-level market shares with complementarity. Technical report, 2020.

Appendix A Proofs of Main Results

A.1 Preliminary Lemmas

Proof of Lemma 1. This follows line by line from the proof of Allen and Rehbeck [2019a], Theorem 1. The statement of that result included an additional Assumption 1, which was not used in the proof as long as the underlying choice is appropriately measurable. Here, we start with Y of the form $Y(X, \beta, \varepsilon)$, which is automatically a measurable function. □

Proof of Lemma 2. See Allen and Rehbeck [2019a], Lemma 1. The result may also be directly proven from Rockafellar [1970], Theorems 23.5 and 25.1. □

Proof of Lemma 3. The function V is convex. The result then follows from Rockafellar [1970], Theorem 4.5, plus repeated differentiation. □

A.2 Proof of Theorem 1

The following lemmas maintain the assumptions of Theorem 1. These assumptions ensure the requisite smoothness assumptions and ensure that the following arguments do not divide by zero.

In order to simplify presentation, we require some additional notation. Let (γ, ξ) be a tuple with $\gamma \in \{1, \dots, K\}^M$ denoting good indices, and let $\xi_k \in \{1, \dots d_{\gamma_k}\}$ describe which characteristic corresponds to the γ_k-th good. We set

$$\partial_{(\gamma,\xi)} \overline{Y}_k(0, \beta) = \partial_{x_{\gamma_1,\xi_1}} \cdots \partial_{x_{\gamma_M,\xi_M}} \overline{Y}_k(0, \beta).$$

As shorthand we also write multiplication of the coefficients of β for the characteristics of (γ, ξ) as

$$\beta_{(\gamma,\xi)} = \beta_{\gamma_1,\xi_1} \cdots \beta_{\gamma_M,\xi_M}.$$

Lemma A.1.

$$\partial_{(\gamma,\xi)} \overline{Y}_k(0, \beta) = \partial_\gamma \partial_k V(0) \beta_{(\gamma,\xi)}.$$

Proof. Lemma 2 establishes

$$\overline{Y}_k(x, \beta) = \partial_k V(\beta_1' x_1, \dots, \beta_K' x_K).$$

Differentiating with respect to x_{γ_1, ξ_1} and evaluating at $x = 0$ yields

$$\partial_{\gamma_1, \xi_1} \overline{Y}_k(0, \beta) = \partial_{\gamma_1} \partial_k V(0) \beta_{\gamma_1, \xi_1}.$$

By repeating the differentiation process and evaluating at $x = 0$ the result follows. Note that in this step, we use the exclusion restriction that the j-th regressors x_j are excluded from the desirability indices of the other goods. □

Lemma A.2.

$$\partial_{(\gamma, \xi)} \overline{Y}_k(0) = \partial_\gamma \partial_k V(0) \int \beta_{(\gamma, \xi)} d\nu(\beta)$$

Proof. We obtain

$$\begin{aligned}
\partial_{(\gamma, \xi)} \overline{Y}_k(0) &= \partial_{x_{1,\xi_1}} \cdots \partial_{x_{M,\xi_M}} \overline{Y}_k(0) \\
&= \int \partial_{x_{1,\xi_1}} \cdots \partial_{x_{M,\xi_M}} \overline{Y}_k(0, \beta) d\nu(\beta) \\
&= \int \partial_\gamma \partial_k V(0) \beta_{(\gamma, \xi)} d\nu(\beta) \\
&= \partial_\gamma \partial_k V(0) \int \beta_{(\gamma, \xi)} d\nu(\beta)
\end{aligned}$$

where the interchange of integration and differentiation in the second equality follows from Assumption 5(i), the third equality is Lemma A.1, and the final equality follows since the evaluation of $\partial_\gamma \partial_k V(0)$ is a constant that does not depend on β. □

Combining the result of Lemma A.2 and Assumption 6 ensures that there exists a set of goods and characteristic indices (γ, ξ) such that $\partial_{(\gamma, \xi)} \overline{Y}_k(0) \neq 0$. To see this, recall that Assumption 6 requires that for each collection of good indices γ we can find characteristic indices ξ such that $\int \beta_{(\gamma, \xi)} d\nu(\beta) \neq 0$. Given Assumption 5(iv) that $\delta_\gamma \delta_k V(0) \neq 0$, Lemma A.2 shows that we must have $\partial_{(\gamma, \xi)} \overline{Y}_k(0) \neq 0$.

Lemma A.3. *Fix $j, k \in \{1, \dots, K\}$ and let $\gamma, \delta \in \{1, K\}^M$. Suppose that each good*

index shows up exactly the same number of times in (γ, k) and (δ, j). Then

$$\partial_\gamma \partial_k V(0) = \partial_\delta \partial_j V(0).$$

Proof. This is a slight restatement of Lemma 3. $\qquad\square$

Suppose now that (δ, η) is defined similar to (γ, ξ). That is, $\delta \in \{1, \dots K\}^M$ denotes good indices and $\eta_j \in \{1, \dots, d_{\delta_j}\}$ indexes a characteristic of the δ_j-th good.

Combining the previous two lemmas, we obtain that if each good index shows up exactly the same number of times in (γ, k) and (δ, j), then

$$\partial_{(\gamma,\xi)} \overline{Y}_k(0) \Big/ \partial_{(\delta,\eta)} \overline{Y}_j(0) = \int \beta_{(\gamma,\xi)} d\nu(\beta) \Big/ \int \beta_{(\delta,\eta)} d\nu(\beta) \qquad (8)$$

whenever the denominator is nonzero. Thus, if the denominator of moments is identified, the numerator is as well.

Lemma A.4. *Suppose $\gamma, \delta \in \{1, \dots, K\}^M$ only differ in at most one component and $\int \beta_{(\delta,\eta)} d\nu(\beta)$ is identified and nonzero. Then for every ξ tuple of characteristic indices $\int \beta_{(\gamma,\xi)} d\nu(\beta)$ is identified.*

Proof. Lemmas A.2 and A.3 immediately imply (8) since

$$\partial_{(\gamma,\xi)} \overline{Y}_k(0) \Big/ \partial_{(\delta,\eta)} \overline{Y}_j(0) = \left(\partial_\gamma \partial_k V(0) \int \beta_{(\gamma,\xi)} d\nu(\beta)\right) \Big/ \left(\partial_\delta \partial_j V(0) \int \beta_{(\delta,\eta)} d\nu(\beta)\right)$$

$$= \int \beta_{(\gamma,\xi)} d\nu(\beta) \Big/ \int \beta_{(\delta,\eta)} d\nu(\beta).$$

The term $\int \beta_{(\gamma,\xi)} d\nu(\beta)$ is identified because all other parts of (8) are identified. $\qquad\square$

Note that in Lemma A.4 that the γ and δ terms can be the same. This covers the non-trivial $K = 1$ case when there are multiple characteristics for the first good.

Theorem 1 requires that $\int \beta_{1,1}^M \nu(d\beta)$ be known. We present a lemma that drops this assumption for the moment. This lemma will be used in subsequent results.

Lemma A.5. *If $\int \beta_{1,1}^M d\nu(\beta)$ is not known, then we still identify the ratio of any M-th order moments*

$$\int \beta_{(\gamma,\xi)} d\nu(\beta) \Big/ \int \beta_{(\delta,\eta)} d\nu(\beta),$$

provided the denominator is nonzero.

Proof. Start with good indices $\delta^0 = (1, \dots, 1)$ of length M, and characteristic indices η^0 such that the corresponding moment of β is nonzero. Applying Lemma A.4 for the pair with goods $\delta^1 = (2, 1, \dots, 1)$ and characteristic indices η^1, we identify the ratio

$$\int \beta_{(\delta^1,\eta^1)} d\nu(\beta) \Big/ \int \beta_{(\delta^0,\eta^0)} d\nu(\beta).$$

We can repeat this procedure with a sequence (δ^1, η^1) and $\delta^2 = (1, 2, \dots, 1)$ with appropriately chosen characteristic indices η^2, and so forth, to construct a sequence $\delta^0, \delta^1, \dots$ that reaches all possible tuples of good indices $\gamma \in \{1, \dots, M\}^K$. At each step, we can change the good index one component at a time and then apply (8). This identifies the ratio of two adjacent moments in this sequence. We avoid dividing by zero because of the relevance condition (Assumption 6), which implies for each set of goods, δ, we can find tuples of characteristics, η, such that $\int \beta_{(\delta,\eta)} d\nu(\beta)$ is nonzero.

By multiplication we can identify new ratios. For example, a ratio involving δ^2 and δ^0 is identified via

$$\int \beta_{(\delta^2,\eta^2)} d\nu(\beta) \Big/ \int \beta_{(\delta^0,\eta^0)} d\nu(\beta) =$$

$$\left(\int \beta_{(\delta^2,\eta^2)} d\nu(\beta) \Big/ \int \beta_{(\delta^1,\eta^1)} d\nu(\beta) \right) \left(\int \beta_{(\delta^1,\eta^1)} d\nu(\beta) \Big/ \int \beta_{(\delta^0,\eta^0)} d\nu(\beta) \right).$$

From these arguments, for each pair of good indices γ and δ we can find some tuples of characteristic indices ξ and η such that

$$\int \beta_{(\gamma,\xi)} d\nu(\beta) \Big/ \int \beta_{(\delta,\eta)} d\nu(\beta)$$

is identified, where numerator and denominator are nonzero.

From Lemma A.4, the ratio

$$\int \beta_{(\delta,\tilde{\eta})} d\nu(\beta) \bigg/ \int \beta_{(\delta,\eta)} d\nu(\beta)$$

is identified for any vector of characteristic indices $\tilde{\eta}$ where η is chosen so that the denominator is nonzero. Thus, we identify the ratio of all moments. $\qquad\square$

Using Lemma A.5, we conclude that if we fix $\int \beta_{1,1}^{M} d\nu(\beta)$ in advance and it is nonzero, we identify all moments. However, we could fix any other nonzero M-th order moment and also obtain identification.

Finally, from Lemma A.2 we have for all $\gamma \in \{1, \ldots, K\}^{M}$ that

$$\partial_{(\gamma,\xi)} \overline{Y}_{k}(0) = \partial_{\gamma} \partial_{k} V(0) \int \beta_{(\gamma,\xi)} d\nu(\beta).$$

Moreover, from Assumption 6 we can find some ξ such that the right hand side is nonzero. By dividing, we identify $\partial_{\gamma} \partial_{k} V(0)$, completing the proof of Theorem 1.

A.3 Proof of Proposition 1

First, from the envelope theorem (see Lemma A.2 above),

$$\frac{\partial \overline{Y}_{1}(0)}{\partial x_{1,1}} = \partial_{1,1} V(0) \int \beta_{1,1} \nu(d\beta).$$

The function V is convex and hence $\partial_{1,1} V(0) > 0$ whenever this derivative is nonzero, so $\frac{\partial \overline{Y}_{1}(0)}{\partial x_{1,1}}$ and $\int \beta_{1,1} d\nu(\beta)$ have the same sign. Thus, the sign of the first moment of $\beta_{1,1}$ is identified from above and the magnitude is assumed known in Assumption 7.

We prove the remainder of the result by induction on M. Recall that with $M = 1$, first order moments are identified from Theorem 1 using the assumption that $\int \beta_{1,1} \nu(d\beta)$ is known and nonzero.

Now, fix an M such that $1 \leqslant M \leqslant \overline{M} - 1$. As the inductive hypothesis, we assume all M-th order moments $\int \beta_{(\delta,\eta)} d\nu(\beta)$ are identified for all $\delta \in \{1, \ldots, K\}^{M}$ and η

collections of characteristic indices. We show all $M + 1$ order moments are also identified.

By Assumption 6, when $K \geqslant 2$, for $\delta \in \{2, \ldots, K\}^M$ (i.e. no good index is equal to 1) we can find a collection of characteristic indices η such that $\int \beta_{(\delta,\eta)} d\nu(\beta) \neq 0$. If instead $K = 1$, we can set δ as the length-M vector of 1's and let η be some collection of characteristic indices with $\eta_m \neq 1$ such that

$$\int \beta_{1,\eta_1} \cdots \beta_{1,\eta_M} d\nu(\beta) \neq 0.$$

In either case $K = 1$ or $K \geqslant 2$, set $\tilde{\delta} = (\delta', 1)'$ and $\tilde{\eta} = (\eta', 1)'$. Then we obtain

$$\int \beta_{(\tilde{\delta},\tilde{\eta})} d\nu(\beta) = \int \beta_{1,1} d\nu(\beta) \int \beta_{(\delta,\eta)} d\nu(\beta)$$

because $\beta_{1,1}$ is independent of all other components of β under the measure ν, and the tuple (δ, η) does not include the first characteristic of good 1. In particular, we identify

$$\int \beta_{(\tilde{\delta},\tilde{\eta})} d\nu(\beta),$$

which is nonzero because it is the product of two nonzero terms. From Lemma A.5, we identify the ratio of all $M + 1$ order moments to $\int \beta_{(\tilde{\delta},\tilde{\eta})} d\nu(\beta)$. Since $\int \beta_{(\tilde{\delta},\tilde{\eta})} d\nu(\beta)$ is known and nonzero we identify all $M + 1$ order moments from Theorem 1. This establishes identification of the moments. To identify derivatives of V, use Lemma A.2 as in the proof of Theorem 1.

A.4   Proof of Proposition 2

Let $x = (x'_1, \ldots, x'_K)'$ satisfy $x_{k,1} \in [\underline{x}_{k,1}, \overline{x}_{k,1}]$ for each k, and $x_{k,j} = 0$ for $j > 1$. Let $x_{:,1} = (x_{1,1}, \ldots, x_{K,1})$ be a vector of the first characteristics for each good. From Lemma 2 and integrating over β, we obtain

$$\overline{Y}(x) = \nabla V(x_{:,1})$$

where $V(x)$ is convex.

Consider initial characteristic values, x^I, and final characteristic values, x^F, such that for all $k \in \{1, \ldots, K\}$ and for all $j > 1$, the equality $x^I_{k,j} = x^F_{k,j} = 0$ holds. By integrating from x^I to x^F, we obtain

$$V(x^F_{:,1}) - V(x^I_{:,1}) = \int_0^1 \overline{Y}(tx^F - (1-t)x^I) \cdot (x^F_{:,1} - x^I_{:,1})dt,$$

where Riemann integrability follows from Rockafellar [1970] Corollary 24.2.1.

Appendix B   Supplemental Results

B.1   Plug-in Estimation

The proof of Theorem 1 is based on multiplying derivative ratios. By directly pluging-in an estimator of the M-th order derivatives, one can construct an estimator of the M-th order moments. Suppose we have estimator

$$\partial_{x_{k_1,\ell_1}} \cdots \partial_{x_{k_M,\ell_M}} \hat{\overline{Y}}_1(0)$$

of the associated M-th order derivative of the average structural function where $k_m = 1$ for all but one term which has $k_{\tilde{m}} = j$. In addition, suppose we have an estimator

$$\partial_{x_{1,1}} \cdots \partial_{x_{1,1}} \hat{\overline{Y}}_j(0)$$

of the M-th order partial derivative of the structural function with respect to the $x_{1,1}$ regressor.

We construct an estimator

$$\int \widehat{\beta_{k_1,\ell_1} \cdots \beta_{k_M,\ell_M}} \nu(d\beta) = \frac{\partial_{x_{k_1,\ell_1}} \cdots \partial_{x_{k_M,\ell_M}} \hat{\overline{Y}}_1(0)}{\partial_{x_{1,1}} \cdots \partial_{x_{1,1}} \hat{\overline{Y}}_j(0)},$$

where for simplicitly we assume $\int \beta_{1,1}^M \nu(d\beta) = 1$. (More generally we need it to be known *a priori* and nonzero.)

Using Equation 8, we see that

$$\int \beta_{k_1,\ell_1} \cdots \beta_{k_M,\ell_M} \nu(d\beta) - \int \widehat{\beta_{k_1,\ell_1} \cdots \beta_{k_M,\ell_M}} \nu(d\beta) =$$

$$\frac{\partial_{x_{k_1,\ell_1}} \cdots \partial_{x_{k_M,\ell_M}} \overline{Y}_1(0)}{\partial_{x_{1,1}} \cdots \partial_{x_{1,1}} \overline{Y}_j(0)} - \frac{\partial_{x_{k_1,\ell_1}} \cdots \partial_{x_{k_M,\ell_M}} \hat{\overline{Y}}_1(0)}{\partial_{x_{1,1}} \cdots \partial_{x_{1,1}} \hat{\overline{Y}}_j(0)}.$$

Thus, estimation error on the right-hand side translates to estimation error on the left hand side for the M-th order moment of β. Note that the choice of the j-th good is arbitrary here, and so one could also construct an estimator with right hand side replaced by an average over ratios with respect to different goods.

This argument can be generalized to additional moments. Here, we use the fact that we are interested in an M-th order moment that is only one good index away from being a vector of 1's. For other M-th order moments, our constructive formulas show one must multiply additional derivative ratios. See in particular the proof of Lemma A.5.

B.2 V Known and Relation to Fox et al. [2012]

Theorem 1 identifies the M-th order moment of β when we fix $\int \beta_{1,1}^M \nu(d\beta)$. By fixing the entire distribution of $\beta_{1,1}$, we can identify all moments of β. An alternative assumption is that V is known. If we impose this assumption, then we can drop Assumption 4, which provides knowledge of each M-th order moment of $\beta_{1,1}$, and Assumption 6 that a rich collection of moments are nonzero. The intuition why we can relax the scale assumption on moments is that here we instead set the scale of V.

Assumption B.1. *V is known in a neighborhood of 0, up to an additive constant.*

Proposition B.1. *Let Assumptions 1-3, 5, and B.1 hold with the same natural number M. Each M-th order moment*

$$\int \beta_{k_1,\ell_1} \cdots \beta_{k_M,\ell_M} d\nu(\beta)$$

is identified.

Proof. Lemma A.2 holds under the assumptions of this proposition, and so for each

tuple of good indices $\gamma \in \{1, \dots, M\}^M$ and characteristic indices ξ we have

$$\partial_{(\gamma,\xi)} \overline{Y}_k(0) = \partial_\gamma \partial_k V(0) \int \beta_{(\gamma,\xi)} d\nu(\beta).$$

Since $\partial_\gamma \partial_k V(0) \neq 0$ by assumption, we identify $\int \beta_{(\gamma,\xi)} \nu(d\beta)$ by

$$\partial_{(\gamma,\xi)} \overline{Y}_k(0) \Big/ \partial_\gamma \partial_k V(0) = \int \beta_{(\gamma,\xi)} d\nu(\beta).$$

$\square$

The proof demonstrates that in fact, we only need $\partial_\gamma \partial_k V(0)$ to be known and nonzero for *some* k in order to identify a corresponding moment of β. For further relation to Theorem 1, consider good indices $\gamma = (1, \dots, 1)$ and characteristic indices $\xi = (1, \dots, 1)$. Then as in the proof of Proposition B.1, we obtain

$$\partial_1^M \overline{Y}_k(0) = \partial_1^M \partial_k V(0) \int \beta_{1,1}^M d\nu(\beta).$$

This shows that one can *either* fix the M-th moment of β_1 or fix $\partial_1^M \partial_k V(0)$ for some k, and then the other can be identified. Thus, Assumption 4 can be replaced in Theorem 1 if we instead assume $\partial_1^M \partial_k V(0)$ is known for some k. Alternatively, if we assume $\beta_{1,1}$ is independent of β as in Section 3.1, then specifying $\partial_{1,1} V(0)$ identifies $\int \beta_{1,1} d\nu(\beta)$ by the envelope theorem

$$\frac{\partial \overline{Y}_1(0)}{\partial x_{1,1}} = \partial_{1,1} V(0) \int \beta_{1,1} d\nu(\beta).$$

Thus, independence combined with a single scale assumption on a partial derivative of V can identify all moments of β by adapting Proposition 1.

In discrete choice, Fox et al. [2012] present a constructive approach to identifying moments of the distribution of random coefficients. Specializing our analysis to discrete choice, their assumptions show that when the distribution of an additive error is known (e.g. logit with known intercept) that this implies identification of V. To see this, recall V is defined as an indirect utility function given a disturbance $\overline{D}$ and constraint set $\overline{B}$. In turn, using Lemma 1 we see $\overline{D}$ and $\overline{B}$ are determined by the

budget set B, disturbance function D, and measure μ over ε. Thus, when the budget set and μ are known then one can find $\overline{B}$ and $\overline{D}$ needed to compute V. For example, multinomial logit is described by

$$V(\vec{u}) = \max_{y \in \overline{B}} \sum_{k=1}^{K} y_k u_k + \sum_{k=1}^{K} (\alpha_k + p_k \ln p_k),$$

where α_k is a nonrandom intercept for good k and $\overline{B}$ is the probability simplex (e.g. Anderson et al. [1992]). The derivatives of V can be used to yield the standard logit formula

$$\overline{Y}_k(\vec{u}) = \frac{e^{\alpha_k + u_k}}{\sum_{j=1}^{K} e^{\alpha_j + u_j}}.$$

We conclude that Proposition B.1 is a generalization of a technique of Fox et al. [2012] to settings outside of discrete choice.[19] However, Theorem 1 does *not* require that V be known. Thus, the results in this paper complement their approach since while it relaxes assumptions on V, it instead requires an additional scale assumption (Assumption 4) and requires that a rich collection of moments are nonzero (Assumption 6).

B.3 Homogeneity of Coefficients and Relation to Chernozhukov et al. [2019a]

When $M = 2$, the proof of Theorem 1 establishes the constructive formula

$$\frac{\partial \overline{Y}_k(0)}{\partial x_{j,\ell}} \Big/ \frac{\partial \overline{Y}_j(0)}{\partial x_{k,m}} = \int \beta_{j,\ell} d\nu(\beta) \Big/ \int \beta_{k,m} d\nu(\beta). \tag{9}$$

A version of (9) has appeared for binary choice in Chernozhukov et al. [2019a], who also discuss identification of the ratios of M-th order moments of β up to scale. They also mention one can identify certain moments up to scale in multinomial choice. Here is one interpretation of their discussion, translated to our setup. Start with a

---

[19]Fox et al. [2012] also presents nonconstructive results using alternative assumptions maintaining the assumption that V is known.

second order derivative like

$$\partial_{x_{k_1,\ell_1}} \partial_{x_{k_2,\ell_2}} \overline{Y}_{k_3}(0) = \partial_{k_1} \partial_{k_2} \partial_{k_3} V(0) \int \beta_{k_1,\ell_1} \beta_{k_2,\ell_2} d\nu(\beta).$$

Now keep the good indices (k_1 and k_2) constant, but change the characteristics to get

$$\partial_{x_{k_1,\tilde{\ell}_1}} \partial_{x_{k_2,\tilde{\ell}_2}} \overline{Y}_{k_3}(0) = \partial_{k_1} \partial_{k_2} \partial_{k_3} V(0) \int \beta_{k_1,\tilde{\ell}_1} \beta_{k_2,\tilde{\ell}_2} d\nu(\beta).$$

Since the derivatives of V are taken with respect to the same arguments, we can divide these equations to identify the associated ratios of moments of β. This technique resembles an implicit function theorem argument for identification. Importantly, this technique only covers ratios of moments in which the good indices (k_1 and k_2 here) are the same, because it does not use symmetry (cf. Lemma 3). Using symmetry, this paper establishes identification of the ratio of *all* M-th order moments, not only those that have the same good indices. However, if we impose additional assumptions such as $\beta_j = \beta_k$ for all goods, then the choice of good indices does not matter. In this special case, using the equations described previously one can identify the ratio of any 2-nd order moments of β. Similar arguments can identify the ratio of any M-th order moments of β.