NONPARAMETRIC IDENTIFICATION OF ENDOGENOUS AND HETEROGENEOUS AGGREGATE DEMAND MODELS: COMPLEMENTS, BUNDLES AND MARKET LEVEL

By

Fabian Dunker, Stefan Hoderlein and Hiroaki Kaido

May 5, 2015

RESEARCH INSTITUTE FOR ECONOMETRICS

DISCUSSION PAPER NO. 4-15

DEPARTMENT OF ECONOMICS
BAR-ILAN UNIVERSITY
RAMAT-GAN 5290002, ISRAEL

http://econ.biu.ac.il/en/node/2473

Nonparametric Identification of Endogenous and Heterogeneous Aggregate Demand Models: Complements, Bundles and the Market Level

Fabian Dunker* Stefan Hoderlein† Hiroaki Kaido‡

University of Goettingen Boston College Boston University

Boston College

June 11, 2014

Abstract

This paper studies nonparametric identification in market level demand models for differentiated products. We generalize common models by allowing for the distribution of heterogeneity parameters (random coefficients) to have a nonparametric distribution across the population and give conditions under which the density of the random coefficients is identified. We show that key identifying restrictions are provided by (i) a set of moment conditions generated by instrumental variables together with an inversion of aggregate demand in unobserved product characteristics; and (ii) an integral transform (Radon transform) that maps the random coefficient density to the aggregate demand. This feature is shown to be common across a wide class of models, and we illustrate this by studying leading demand models. Our examples include demand models based on the multinomial choice (Berry, Levinsohn, Pakes, 1995), the choice of bundles of goods that can be substitutes or complements, and the choice of goods consumed in multiple units.

*Institute for Numerical and Applied Mathematics, University of Goettingen, Lotzestr. 16-18, D-37083 Goettingen, Germany; Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, Email: dunker@math.uni-goettingen.de

†Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, Email: stefan_hoderlein@yahoo.com.

‡Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA, Email: hkaido@bu.edu. We thank Steve Berry, Jeremy Fox, Amit Gandhi, Phil Haile, and Arthur Lewbel for their helpful comments. We also thank conference participants at the BU-BC joint mini-conference, the Demand Estimation and Modeling Conference 2013 and the Cowles Summer Conference 2014. Dunker gratefully acknowledges financial support from DFG FOR916. Kaido gratefully acknowledges financial support from NSF Grant SES-1357643.

1 Introduction

Modeling consumer demand for products that are bought in single or discrete units has a long and colorful history in applied Economics, dating back to at least the foundational work of McFadden (1974, 1981). While allowing for heterogeneity, much of the earlier work on this topic, however, was not able to deal with the fact that in particular the own price is endogenous. In a seminal paper that provides the foundation for much of contemporaneous work on discrete choice consumer demand, Berry, Levinsohn and Pakes (1994, BLP) have proposed a solution to the endogeneity problem. Indeed, this work is so appealing that it is not just applied in discrete choice demand and empirical IO, but also increasingly in many adjacent fields, such as health and urban economics, education and many others. From a methodological perspective, this line of work is quite different from traditional multivariate choice, as it uses data on the aggregate level and integrates out individual characteristics[1] to obtain a system of nonseparable equations. This system is then inverted for unobservables for which in turn a moment condition is then supposed to hold.

Descending in parts from the parametric work of McFadden (1974, 1981), BLP share many of its features, in particular (parametric) distributional assumptions, but also a linear random coefficients (RCs) structure for the latent utility. Not surprisingly, there is increasing interest in the properties of the model, in particular which features of the model are nonparametrically point identified, and how the structural assumptions affect identification of the parameters of interest. Why is the answer to these questions important? Because an empricist working with this model wants to understand whether the results she obtained are a consequence of the specific parametric assumptions she invoked, or whether they are at least qualitatively robust. In addition, nonparametric identification provides some guidance on essential model structure and on data requirements, in particular about instruments. Finally, understanding the basic structure of the model makes it easier to understand how the model can be extended. Extensions of the BLP framework that are desirable are in particular to allow for consumption of bundles and multiple units of a product without modeling every choice as a new separate alternative.

We are not the first to ask the nonparametric identification question for market demand models. In a series of elegant papers, Berry and Haile (2011, 2013, BH henceforth) provide important answers to many of the identification questions. In particular, they establish conditions under which the "Berry inversion", a core building block of the BLP model named after Berry (1994), which allows to solve for unobserved product characteristics, as well as the distribution of a heterogeneous utility index are nonparametrically identified.

Our work complements this line of work in that we follow more closely the original BLP

[1]There are extensions of the BLP framework that allow for the use of Microdata, see Berry, Levinsohn and Pakes (2004, MicroBLP). In this paper, we focus on the aggregate demand version of BLP, and leave an analogous work to MicroBLP for future research.

specification and assume in addition that the utility index has a linear random coefficients (RCs) structure. More specifically, we show how to nonparametrically identify the distribution of random coefficients in this framework. This result does not just close the remaining gap in the proof of nonparametric identification of the original BLP model, but is also important for applications because the distribution of random coefficients allows to characterize the distribution of the changes in welfare due to a change in regressors, in particular the own price (to borrow an analogy from the treatment effect literature, if we think of a price as a treatment, BH recover the treatment effect on the distribution, while we recover the distribution of treatment effects). The arguments in establishing nonparametric identification of these changes are constructive and permit the construction of sample counterparts estimators, using theory in Hoderlein, Klemelä and Mammen (2010). From this theory it is well known that these estimators reveal that the random coefficients density is only weakly identified, suggesting that numerical instabilities and problems frequently reported and discussed in the BLP literature, e.g., Dube, Fox and Su (2013), are caused or aggravated by this feature of the model.

The second contribution in this paper is that we use the insights obtained from the identification results to extend the market demand framework to cover bundle choice (i.e., consume complementary goods together), as well as consumption of multiple units. Note that bundles and multiple purchases can in principle be accommodated within the BLP framework by treating them as separate alternatives. However, this is not parsimonious as the number of alternatives increases rapidly and with it the number of unobserved product characteristics, making the system quickly intractable. To fix ideas, suppose there were two goods, say good A and B. First, we allow for the joint consumption of goods A and B, and second, we allow for the consumption of several units of either A and/or B, without labeling it a separate alternative. We model the utility of each bundle as a combination of the utilities for each good and an extra utility from consuming the bundle. This structure in turn implies that the dimension of the unobservable product characteristic equals the number of goods J instead of the number of bundles. There are three conclusions we draw from this contribution: first, depending on the type of model, the data requirements vary. In particular, to identify all structural parts of the model, in, say, the model on bundle choice, market shares are not the correct dependent variable any more. Second, depending on the object of interest, the data requirements and assumptions may vary depending on whether we want to just recover demand elasticities, or the entire distribution of random coefficients. Third, the parsimonious features of the structural model result in significant overidentification of the model, which opens up the way for specification testing, and efficient estimation. As in the classical BLP setup, in all setups we may use the identification argument to propose a nonparametric sample counterpart estimators, but we also use the insights obtained to propose a parametric estimator for models where there had not been an estimator before.

Related literature: as discussed above, this paper is closely related to both the original

BLP line of work (Berry, Levinsohn and Pakes (1994, 2004)), as well as to the recent identification analysis of Berry and Haile (2011, 2013). Because of its generality, our approach also provides identification analysis for the "pure characteristics" model of Berry and Pakes (2007), see also Ackerberg, Benkard, Berry and Pakes (2007) for an overview. Other important work in this literature that is completely or partially covered by the identification results in this paper include Petrin (2002) and Nevo (2001). Moreover, from a methodological perspective, we note that BLP continues a line of work that emanates from a broader literature which in turn was pioneered by McFadden (1974, 1981); some of our identification results extend therefore beyond the specific market demand model at hand. Other important recent contributions in discrete choice demand include Armstrong (2013) and Moon, Shum, and Weidner (2013).

In addition to this line of work, we also share some commonalities with the work on bundle choice in IO, most notably Gentzkow (2007), and Fox and Lazzati (2013). For some of the examples discussed in this paper, we use Gale-Nikaido inversion results, which are related to arguments in Berry, Gandhi and Haile (2013). Because of the GMM type endogeneity, our approach also relates to nonparametric IV, in particular to Newey and Powell (2003), Andrews (2011), and Dunker, Florens, Hohage, Johannes, and Mammen (2014). Finally, our arguments are related to the literature on random coefficients in discrete choice model, see Ichimura and Thompson (1995), Gautier and Kitamura (2013), Dunker, Hoderlein and Kaido (2013), Fox and Gandhi (2012), and Matzkin (2012). Since we use the Radon transform introduced by Hoderlein, Klemelä and Mammen (2010, HKM) into Econometrics, possibly in conjunction with tensor products as in Dunker, Hoderlein and Kaido (2013), this work is particularly close to the literature that uses the Radon transform, in particular HKM and Gautier and Hoderlein (2013).

Structure of the paper: We introduce the model in the next section, including all definitions and assumptions in the various setups. We then provide general identification results which we specify to the individual models. Because of the difference in requirements, this section is divided into identifying demand elasticities and identifying the entire distribution of random coefficients. Moreover, we discuss each result for each of the specifications we consider. Finally, an outlook concludes.

2 Nonparametric Identification

2.1 Set up

We begin with a setting where a consumer faces $J \in \mathbb{N}$ products and an outside good which is labeled good 0. Throughout, we index products by j and markets by t. We use upper-case letters, e.g. X_{jt}, for random variables (or vectors) that vary across markets and lower-case letters, e.g. x_j, for particular values the random variables (vectors) can take. In addition, we

use letters without a subscript for products e.g. X_t to represent vectors e.g. $(X_{1t}, \cdots, X_{Jt})$. For each individual in market t, the (indirect) utility from consuming good j depends on its (log) price P_{jt}, a vector of observable characteristics $X_{jt} \in \mathbb{R}^k$, and an unobservable scalar characteristic $\Xi_{jt} \in \mathbb{R}$. Following Berry, Levinsohn and Pakes (1995), we model the utility from consuming good j using the linear random coefficient specification:

$$U^*_{jt} \equiv X'_{jt}\beta_j + \alpha_j P_{jt} + \Xi_{jt},\ j = 1, \cdots, J\ , \tag{2.1}$$

where $(\alpha_j, \beta_j) \in \mathbb{R}^{k+1}$ is a vector of random coefficients, which varies across individuals. For notational simplicity, we do not add subscripts for individuals to the random coefficients, but the assumption that (α_j, β_j) (and hence the utility) may vary across individuals should be implicitly understood. Throughout, we assume that X_{jt} is exogenous, while P_{jt} can be correlated with the unobserved product characteristic Ξ_{jt} in an arbitrary way. Without loss of generality, we normalize the utility from the outside good to 0. This mirrors the setup considered in BH (2013).

Note that in order to reduce notational complexity we suppress an individual specific index (usually, i). This is justified as we mainly talk about identification, and we would think of a large sample of individuals as *iid* copies of this population model. The drawback of such an approach is that we have to be very explicit about which variables vary across individuals or markets. The random coefficients for instance vary across individuals in any given market (or, alternatively, have a distribution in any given market in the population), while the product characteristics vary solely across markets. We can in principle allow for all random parameters (α, β) to be individual, alternative and market specific as long as their distribution does not vary across markets (this includes the commonly employed additive "logit error", typically denoted ϵ_{ijt} in the sample, which is *iid* across i, j and t). We can allow for the coefficients to be alternative j specific, and will indeed do so below. However, parts of the analysis will subsequently change, and we start out with the more common case where the coefficients are the same across j.

Having specified the model on individual level, the outcomes of individual decisions are then aggregated in every market. The econometrician observes exactly these market level outcomes $S_{l,t}$, where l belongs to some index set denoted by $\mathbb{L}$. Below, we give the three main settings we consider. The first example is the setting of BLP, where individuals choose a single good out of J products, where J can be any natural number. The second and third examples consider choices among bundles generated from combining two goods. They are settings commonly considered in the literature; see e.g., Gentzkow (2007) and references therein.[2] Throughout the following examples, the random coefficient vector is denoted by $\theta \in \mathbb{R}^{d_\theta}$ and is assumed to have

[2]In this paper, we focus on the bundle choice with two goods. We conjecture that bundles generated from more than two goods could potentially be handled with some additional structure, which we leave as future work.

a joint probability density function f_θ with respect to Lebesgue measure, i.e., be continuously distributed.

Example 1 (BLP)**.** Each individual chooses the product that maximizes her utility out of $J \in \mathbb{N}$ products. Hence, product j is chosen if

$$U^*_{jt} > U^*_{it} \ , \quad \forall i \neq j \ . \tag{2.2}$$

The demand for good j in market t is obtained by aggregating the individual demand with respect to the distribution of individual preferences. For example, if $(\alpha_j, \beta_j) = (\alpha, \beta)$ for all j, *a.s.*, the aggregate demand is given by

$$\begin{aligned}\varphi_j(X_t, P_t, \Xi_t) = \int 1\{X'_{jt}b + aP_{jt} > -\Xi_{jt}\}1\{(X_{jt} - X_{1t})'b + a(P_{jt} - P_{1t}) > -(\Xi_{jt} - \Xi_{1t})\} \\ \cdots 1\{(X_{jt} - X_{Jt})'b + a(P_{jt} - P_{Jt}) > -(\Xi_{jt} - \Xi_{Jt})\} f_\theta(b, a) d\theta \ , \end{aligned} \tag{2.3}$$

for $j = 1, \cdots, J$, while the aggregate demand for good 0 is given by

$$\varphi_0(X_t, P_t, \Xi_t) = \int 1\{X'_{1t}b + aP_{1t} < -\Xi_{1t}\} \cdots 1\{X'_{Jt}b + aP_{Jt} < -\Xi_{Jt}\} f_\theta(b, a) d\theta \ . \tag{2.4}$$

The researcher then observes the market shares of products $S_{lt} = \varphi_l(X_t, P_t, \Xi_t), l \in \mathbb{L}$, where $\mathbb{L} = \{0, 1, \cdots, J\}$.

Example 2 (Bundles)**.** Each individual faces $J = 2$ products and decides whether or not to consume a single unit of each of the products. There are therefore four possible combinations (Y_1, Y_2) of consumption units, which we call *bundles*. In addition to the utility from consuming each good as in (2.1), the individuals gain additional utility (or disutility) Δ if the two goods are consumed simultaneously. Here, Δ is also allowed to vary across individuals. The utility $U^*_{(Y_1,Y_2),t}$ from each bundle is therefore specified as follows:

$$\begin{aligned} &U^*_{(0,0),t} = 0, \quad U^*_{(1,0),t} = X'_{1t}\beta_1 + \alpha_1 P_{1t} + \Xi_{1t} \\ &U^*_{(0,1),t} = X'_{2t}\beta_2 + \alpha_2 P_{2t} + \Xi_{2t}, \quad U^*_{(1,1),t} = X'_{1t}\beta_1 + X'_{2t}\beta_2 + \alpha_1 P_{1t} + \alpha_2 P_{2t} + \Delta + \Xi_{1t} + \Xi_{2t} \ , \end{aligned} \tag{2.5}$$

Each individual chooses a bundle that maximizes her utility. Hence, bundle (y_1, y_2) is chosen when $U^*_{(y_1,y_2),t} > U^*_{(y'_1,y'_2),t}$ for all $(y'_1, y'_2) \neq (y_1, y_2)$. For example, assuming $(\beta_j, \alpha_j) = (\beta, \alpha)$ for all j, bundle $(1, 0)$ is chosen if

$$\begin{aligned} X'_{1t}\beta + \alpha P_{1t} + \Xi_{1t} > 0, \quad \text{and} \quad X'_{1t}\beta + \alpha P_{1t} + \Xi_{1t} > X'_{2t}\beta + \alpha P_{2t} + \Xi_{2t}, \quad \text{and} \\ X'_{1t}\beta + \alpha P_{1t} + \xi_{1t} > X'_{1t}\beta + \alpha P_{1t} + \Xi_{1t} + X'_{2t}\beta + \alpha P_{2t} + \Xi_{2t} + \Delta \ . \end{aligned} \tag{2.6}$$

Suppose the random coefficients $\theta = (\beta', \alpha, \Delta)$ have a joint density f_θ. The aggregate structural demand for $(1,0)$ can then be obtained by integrating over the set of individuals satisfying (2.6) with respect to the distribution of the random coefficients:

$$\varphi_{(1,0)}(X_t, P_t, \Xi_t) = \int 1\{X_{1t}'b + aP_{1t} > -\Xi_{1t}\} 1\{(X_{1t} - X_{2t})'b + a(P_{1t} - P_{2t}) > \Xi_{2t} - \Xi_{1t}\} \\ \times 1\{X_{2t}'b + aP_{2t} + \Delta < -\Xi_{2t}\} f_\theta(b, a, \Delta) d\theta \ . \quad (2.7)$$

The aggregate demand on other bundles can be obtained similarly. The econometrician then observes a vector of aggregate demand on the bundles: $S_{l,t} = \varphi_l(X_t, P_t, \Xi_t), l \in \mathbb{L}$ where $\mathbb{L} \equiv \{(0,0), (1,0), (0,1), (1,1)\}$.

Example 3 (Multiple Units of Consumption)**.** Each individual faces $J = 2$ products, but she may consume multiple units of each product. For simplicity, we consider the simplest setup where $Y_1 \in \{0,1,2\}$ and $Y_2 \in \{0,1\}$. The utility from consuming y_1 units of product 1 and y_2 units of product 2 is specified as follows:

$$U^*_{(y_1,y_2),t} = y_1 U^*_{1t} + y_2 U^*_{2t} + \Delta_{(y_1,y_2)} \ , \quad (2.8)$$

where $\Delta_{(y_1,y_2)}$ is the additional utility (or disutility) from consuming the particular bundle (y_1, y_2). This specification allows, e.g., for decreasing marginal utility (with the number of units), as well as interaction effects. We assume that $\Delta_{(1,0)} = \Delta_{(0,1)} = 0$ as U^*_{1t} and U^*_{2t} give the utility from consuming a single unit of each of the two goods. Throughout this example, we assume that $U^*_{(y_1,y_2),t}$ is concave in (y_1, y_2). Then, a bundle is chosen if its utility exceeds those of the neighboring alternatives. For example, assuming $(\beta_j, \alpha_j) = (\beta, \alpha)$ for all j, bundle $(2,0)$ is chosen if it is preferred to bundles (1,0), (1,1) and (2,1). That is,

$$\begin{aligned} &2(X_{1t}'\beta + \alpha P_{1t} + \Xi_{1t}) + \Delta_{(2,0)} > X_{1t}'\beta + \alpha P_{1t} + \Xi_{1t} \ , \\ &\quad 2(X_{1t}'\beta + \alpha P_{1t} + \Xi_{1t}) + \Delta_{(2,0)} > X_{1t}'\beta + \alpha P_{1t} + \Xi_{1t} + X_{2t}'\beta + \alpha P_{2t} + \Xi_{2t} + \Delta_{(1,1)} \\ &\text{and} \ \ 2(X_{1t}'\beta + \alpha P_{1t} + \Xi_{1t}) + \Delta_{(2,0)} > 2(X_{1t}'\beta + \alpha P_{1t} + \Xi_{1t}) + X_{2t}'\beta + \alpha P_{2t} + \Xi_{2t} + \Delta_{(2,1)}. \end{aligned} \quad (2.9)$$

The aggregate structural demand can be obtained as

$$\varphi_{(2,0)}(X_t, P_t, \Xi_t) = \int 1\{X_{1t}'b + aP_{1t} + \Delta_{(2,0)} > -\Xi_{1t}\} \\ \times 1\{(X_{1t} - X_{2t})'b + a(P_{1t} - P_{2t}) + \Delta_{(2,0)} - \Delta_{(1,1)} > -\Xi_{1t} + \Xi_{2t}\} \\ \times 1\{X_{2t}'b + aP_{2t} + \Delta_{(2,1)} - \Delta_{(2,0)} < -\Xi_{2t}\} f_\theta(b, a) d\theta \ . \quad (2.10)$$

The observed aggregate demand on the bundles are similarly defined for $S_{l,t} = \varphi_l(X_t, P_t, \Xi_t), l \in$

$\mathbb{L}$ where $\mathbb{L} \equiv \{(0,0),(1,0),(0,1),(1,1),(2,0),(2,1)\}$.

In Examples 2 and 3, we assume that the econometrician observes the aggregate demand for all the respective bundles. We emphasize this point as it changes the data requirement, and an interesting open question arises about what happens if these requirements are not met. Example of data sets that would satisfy these requirements are when 1 individual observations are collected through direct survey or scanner data on individual consumption (in every market), 2. aggregate variables (market shares) are collected, but augmented with a survey that asks individuals whether they consume each good separately or as a bundle. 3. Finally, another possible data source are producer's direct record of sales of bundles, provided each bundles are recorded separately (e.g., when they are sold through promotional activities). When discussing examples 2 and 3, we henceforth tacitly assume to have access to such data in principle.

2.2 Inversion and Identification of Structural Demand

The first step toward identification of f_θ is to use a set of moment conditions generated by instrumental variables to identify the aggregate demand function φ. Following BH (2013), we partition the covariates as $X_{jt} = (X_{jt}^{(1)}, X_{jt}^{(2)}) \in \mathbb{R} \times \mathbb{R}^{k_2}$, where $k_2 = k - 1$ and make the following assumption.

Assumption 2.1. *(i) The coefficient $\beta_j^{(1)}$ on $X_{jt}^{(1)}$ is non-random for all j and is normalized to 1; (ii) $(\beta_j, \alpha_j) = (\beta, \alpha)$ for all j, a.s.*

Assumption 2.1 (i) requires that at least one coefficient on the covariates is non-random. Since we may freely choose the scale of utility, we normalize the utility by setting $\beta_j^{(1)} = 1$ for all j. Assumption 2.1 (ii) requires that the random coefficients (β_j, α_j) are common across alternatives. Requiring that the coefficients on X_{jt} is common is standard in the literature. Assumption 2.1 (ii), however, excludes an alternative specific additive error. Hence, this setting falls in the scope of the pure characteristic model studied in Berry and Pakes (2007). It is possible to relax this assumption. We will discuss this extension in Section 3.1. Under Assumption 2.1, the utility for product j can be written as $U_{jt}^* = X_{jt}^{(2)\prime(2)} + \alpha P_{jt} + D_{jt}$, where $D_{jt} \equiv X_{jt}^{(1)} + \Xi_{jt}$ is the part of the utility that is common across individuals. Assumption 2.1 (i) is arguably strong but will provide a way to obtain valid instruments required to identify the structural demand (see BH, 2013, Section 7 for details). Under this assumption, U_{jt}^* is strictly increasing in D_{jt} but unaffected by D_{it} for all $i \neq j$. In Example 1, together with a mild regularity condition, this is sufficient for inverting the demand system to obtain Ξ_t as a function of the market shares S_t, price P_t, and exogenous covariates X_t (Berry, Gandhi, and Haile, 2013). In Examples 2 and 3, one may also obtain Ξ_t as a function of (S_t, P_t, X_t) in a similar manner. Since we observe aggregate demand for more than two distinct bundles and have two unknowns (Ξ_{1t}, Ξ_{2t}), we may use a subsystem of the demand equations to obtain an inversion.

We therefore make the following high-level assumption. For this, we redefine the aggregate demand as a function of $(X_t^{(2)}, P_t, D_t)$ instead of (X_t, P_t, Ξ_t) by $\phi(X_t^{(2)}, P_t, D_t) \equiv \varphi(X_t, P_t, \Xi_t)$, where $X_t = (X_t^{(1)}, X_t^{(2)})$ and $D_t = \Xi_t + X_t^{(1)}$.

Assumption 2.2. *For some subset $\tilde{\mathbb{L}}$ of $\mathbb{L}$ whose cardinality is J, there exists a unique function $\psi : \mathbb{R}^{Jk_2} \times \mathbb{R}^J \times \mathbb{R}^J \to \mathbb{R}^J$ such that $D_{jt} = \psi_j(X_t^{(2)}, P_t, \tilde{S}_t)$ for $j = 1, \cdots, J$, where $\tilde{S}_t$ is a subvector of S_t, which stacks the components of S_t whose indices belong to $\tilde{\mathbb{L}}$.*

Under Assumption 2.2, we may write

$$\Xi_{jt} = \psi_j(X_t^{(2)}, P_t, \tilde{S}_t) - X_{jt}^{(1)}. \tag{2.11}$$

This can be used to generate moment conditions in order to identify the aggregate demand function.

Example 1 (BLP, continued)**.** Let $\tilde{\mathbb{L}} = \{1, \cdots, J\}$. In this setting, the inversion discussed above is the standard Berry inversion of demand. A key condition for the inversion is that the products are *connected substitutes* (Berry, Gandhi, and Haile (2013)). The linear random coefficient specification as in (2.1) is known to satisfy this condition. Then, Assumption 2.2 follows.

For the next two examples, we consider an alternative procedure (Lemma 1 given in the appendix) for inverting the demand. This is because these examples share a specific structure. We note that the inversion of Berry, Gandhi, and Haile (2013) can still be applied to bundles if one treats each bundle as a separate good and recast the bundle choice problem into a standard multinomial choice problem. However, as can be seen from (2.5) and (2.8), Examples 2 and 3 have the additional structure that the utility of a bundle is the combination of the utilities for each good and extra utilities, and hence the model does not involve any bundle specific unobserved characteristic. This structure in turn implies that the dimension of the unobservable product characteristic Ξ_t equals the number of goods J, while the econometrician observes $dim(S) = \Pi_{j=1}^J (d_j + 1)$ aggregate choice probabilities over bundles, where d_j is the maximum number of consumption units allowed for each good (e.g. in Example 2, $J = 2$, and $dim(S) = 4$). This suggests that (i) using only a part of the demand system is sufficient for obtaining an inversion, which can be used to identify f_θ and (ii) using additional subcomponents of S, one may potentially overidentify the parameter of interest. We therefore consider an inversion that exploits a monotonicity property of the demand system that follows from this structure.[3] For this, we assume that the following condition is met.

[3]The additional structure can potentially be tested. For example, in Example 2, one may identify the demand for bundles (1,0) and (1,1) using the inversion described below under the hypothesis that eq. (2.5) holds. Further, treating (1,0), (0,1), and (1,1) as three separate goods (and (0,0) as an outside good) and applying the inversion of Berry, Gandhi, and Haile (2013), one may identify the demand for bundles (1,0) and (1,1) without imposing (2.5). The specification can then be tested by comparing the demand functions obtained from these distinct inversions. We are indebted to Phil Haile for this point.

Condition 2.1. *The random coefficient density f_θ is continuously differentiable. $(\beta^{(2)}, \alpha)$ and D have full support in $\mathbb{R}^{k_2+1}$ and $\mathbb{R}^2$ respectively.*

Example 2 (Bundles, continued)**.** Let $\tilde{\mathbb{L}} = \{(1,0),(1,1)\}$. From (2.7), it is straightforward to show that $\varphi_{(1,0)}$ is increasing in D_1 but is decreasing in D_2, while $\varphi_{(1,1)}$ is increasing both in D_1 and D_2. Hence, the Jacobian matrix is non-degenerate. Together with a mild support condition on (D_1, D_2), this allows to invert the demand (sub)system and write $\Xi_j = \psi_j(X_t^{(2)}, P_t, \tilde{S}_t) - X_{jt}^{(1)}$, where $\tilde{S}_t = (S_{(1,0),t}, S_{(1,1),t})$. One may alternatively choose $\tilde{\mathbb{L}} = \{(0,0),(0,1)\}$, and the argument is similar. This inversion is valid even if the two goods are complements. This is because the inversion uses the monotonicity property of the aggregate choice probabilities on bundles (e.g. $\phi_{(1,0)}$ and $\phi_{(1,1)}$) with respect to (D_1, D_2). Hence, even if the aggregate share of each good (e.g. aggregate share on good 1: $\sigma_1 = \phi_{(1,0)} + \phi_{(1,1)}$) is not invertible in the price P_t due to the presence of complementary goods, one can still obtain a useful inversion provided that aggregate choice probabilities on bundles are observed.

Example 3 (Multiple units, continued)**.** Let $\tilde{\mathbb{L}} = \{(2,0),(2,1)\}$. From (2.9), $\varphi_{(2,0)}$ is increasing in D_1 but is decreasing in D_2. Similarly, $\varphi_{(2,1)}$ is increasing in both D_1 and D_2. The rest of the argument is similar to Example 2. One may alternatively take $\tilde{\mathbb{L}} = \{(0,0),(0,1)\}$ and use the same line of argument. Note, however, that (1,0) or (1,1) cannot be included in $\tilde{\mathbb{L}}$ as $\phi_{(1,0)}$ and $\phi_{(1,1)}$ are not monotonic in one of (D_1, D_2). This is because increasing D_1 while fixing D_2, for example, makes good 1 more attractive and creates both an inflow of individuals who move from (0,0) to (1,0) and an outflow of individuals who move from (1,0) to (2,0). Hence, the demand for (1,0) does not necessarily change monotonically.

The inverted system in (2.11), together with the following assumption, yields a set of moment conditions the researcher can use to identify the structural demand.

Assumption 2.3. *There is a vector of instrumental variables $Z_t \in \mathbb{R}^{d_Z}$ such that (i) $E[\Xi_{jt}|Z_t, X_t] = 0, a.s.$; (ii) for any $B : \mathbb{R}^{Jk_2} \times \mathbb{R}^J \times \mathbb{R}^J \to \mathbb{R}$ with $E[|B(X_t^{(2)}, P_t, \tilde{S}_t)|] < \infty$, it holds that*

$$E[B(X_t^{(2)}, P_t, \tilde{S}_t)|Z_t, X_t] = 0 \Longrightarrow B(X_t^{(2)}, P_t, \tilde{S}_t) = 0, \ a.s.$$

Assumption 2.3 (i) is a mean independence assumption on Ξ_{jt} given a set of instruments Z_t, which also normalizes the location of Ξ_{jt}. Assumption 2.3 (ii) is a completeness condition, which is common in the nonparametric IV literature, see BH (2013) for a detailed discussion. However, the role it plays here is slightly different, as the moment condition leads to an integral equation which is different from nonparametric IV (Newey & Powell, 2003), and more resembles GMM. As such, the construction of a sample counterpart estimator is less clear. In Section 3.2, we discuss an approach based on a strengthening of the mean independence condition to full independence. In case such a strengthening is economically palatable, we still retain the sum $X_{jt}^{(1)} + \Xi_{jt}$, which has a closer analogy to a dependent variable in nonparametric IV.

Given Assumption 2.3 and (2.11), the unknown function ψ can be identified through the following conditional moment restrictions:

$$E[\psi_j(X_t^{(2)}, P_t, S_t) - X_{jt}^{(1)} | Z_t, X_t] = 0, \quad j = 1, \cdots, J. \tag{2.12}$$

We here state this result as a theorem.

Theorem 2.2. *Suppose Assumptions 2.1-2.3 hold. Then, ψ is identified.*

Once ψ is identified, the structural demand ϕ can be identified nonparametrically in Examples 1 and 2.

Example 1 (BLP, continued)**.** Recall that ψ is a unique function such that

$$S_{jt} = \phi_j(X_t^{(2)}, P_t, D_t),\ j = 1, \cdots, J \quad \Leftrightarrow \quad \Xi_{jt} = \psi_j(X_t^{(2)}, P_t, \tilde{S}_t) - X_{jt}^{(1)}, \quad j = 1, \cdots, J, \tag{2.13}$$

where $\tilde{S}_t = (S_{1t}, \cdots, S_{Jt})$. Hence, the structural demand $(\phi_1, \cdots, \phi_J)$ is identified by Theorem 2.2 and the equivalence relation above. In addition, ϕ_0 is identified through the identity: $\phi_0 = 1 - \sum_{j=1}^{J} \phi_j$.

Example 2 (Bundles, continued)**.** Let $\tilde{\mathbb{L}} = \{(1,0), (1,1)\}$ as before. ψ is then a unique function such that

$$S_{lt} = \phi_l(X_t^{(2)}, P_t, D_t),\ l \in \tilde{\mathbb{L}} \quad \Leftrightarrow \quad \Xi_{jt} = \psi_j(X_t^{(2)}, P_t, \tilde{S}_t) - X_t^{(1)}, \quad j = 1, 2, \tag{2.14}$$

where $\tilde{S}_t = (S_{(1,0),t}, S_{(1,1),t})$. Theorem 2.2 and the equivalence relation above then identify the demand for bundles $(1,0)$ and $(1,1)$. This, therefore, only identifies subcomponents of ϕ. Although these subcomponents are sufficient for recovering the random coefficient density as we will show in the next section, one may also identify the rest of the subcomponents by taking $\tilde{\mathbb{L}} = \{(0,0), (0,1)\}$ and applying Theorem 2.2 again.

In Example 3, identification of ψ allows to identify key components of ϕ but not the entire vector.

Example 3 (Multiple units, continued)**.** Analogous to Example 2, ψ is a unique function that satisfies (2.14) with $\tilde{\mathbb{L}} = \{(2,0), (2,1)\}$. Theorem 2.2 and the equivalence relation in (2.14) then identify the demand for the bundles as in the last example. Similarly, one may identify the demand for bundles $(0,0)$ and $(0,1)$ by taking $\tilde{\mathbb{L}} = \{(0,0), (0,1)\}$. However, the demand for the middle alternatives $(1,0)$ and $(1,1)$ are not directly identified due to their lack of monotonicity in (D_1, D_2). We will revisit identification of the demand for these alternatives in the next section.

2.3 Identification of the Random Coefficient Density

Our goal is to identify the random coefficient density f_θ. Identification of this object is particularly important for welfare analysis. For example, consider a change in the characteristic of good j from $(x_j^{(2)}, p_j, \delta_j)$ to $(\tilde{x}_j^{(2)}, \tilde{p}_j, \tilde{\delta}_j)$. The change may be due to a new regulation, an improvement of the quality of a product, or an introduction of a new product. Knowledge of f_θ would allow the researcher to calculate the distribution of the welfare change: $(\tilde{x}_j^{(2)} - x_j^{(2)})'^{(2)} + \alpha(\tilde{p}_j - p_j) + (\tilde{\delta}_j - \delta_j)$. This, for example, allows to predict the fraction of consumers who would benefit from the change and therefore prefers the product with a new characteristic.

Our strategy for identification of the random coefficient density is to construct a function from the structural demand, which is related to the density through an integral transform known as the *Radon transform*. More precisely, for some $l \in \mathbb{L}$, we construct a function $\Phi_l(w, u)$ such that

$$\frac{\partial \Phi_l(w,u)}{\partial u} = \mathcal{R}[f_\theta](w,u) \ , \tag{2.15}$$

where $w \in \mathbb{S}^{d_\theta - 1}$ is a unit vector that depends on $(x_j^{(2)}, p_j)$ for some $j \in \{1, \cdots, J\}$, and $u \in \mathbb{R}$ is a scalar that depends on $(x_j^{(2)}, p_j, \delta_j)$. $\mathcal{R}$ is the Radon transform defined pointwise by

$$\mathcal{R}[f](w,u) = \int_{P_{w,u}} f(v) d\mu_{w,u}(v). \tag{2.16}$$

where $P_{w,u}$ denotes the hyperplane $\{v \in \mathbb{R}^{d_\theta} : v'w = u\}$, and $\mu_{w,u}$ is the Lebesgue measure on $P_{w,u}$. (See for example Helgason (1999) for details on the properties of the Radon transform including its injectivity.) Our identification strategy is constructive and will therefore suggest a natural nonparametric estimator. Applications of the Radon transform to random coefficients models have been studied in Hoderlein, Klemelä, and Mammen (2010), and Gautier and Hoderlein (2013).

Let $\mathcal{S}(\mathbb{R}^{d_\theta})$ be the space of rapidly decreasing functions (Schwartz space) on $\mathbb{R}^{d_\theta}$. Throughout, we maintain the following assumption.

Assumption 2.4. *(i) For all $i \in \{1, \cdots, J\}$, $(X_{it}^{(2)}, P_{it}, D_{it})$ are absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^{k_2+2}$; (ii) For some $j \in \{1, \cdots, J\}$, $(X_{jt}^{(2)}, P_{jt}, D_{jt})$ has a full support; (iii) the random coefficients θ are independent of (X_t, P_t, D_t), and its density f_θ is in $\mathcal{S}(\mathbb{R}^{d_\theta})$.*

Assumption 2.4 (i) requires that $(X_{it}^{(2)}, P_{it}, D_{it})$ are continuously distributed for all i. Assumption 2.4 (ii) then requires that for at least one product, $(X_{jt}^{(2)}, P_{jt})$ has full support, and by Assumption 2.4 (iii), we assume that the covariates (X_t, P_t, D_t) are exogenous to the individual heterogeneity. These conditions are used to invert the Radon transform in (2.16). The construction of Φ_l requires slightly different arguments and sets of assumptions across Examples

1-3. We therefore discuss them separately.

Example 1 (BLP, continued)**.** Recall that the demand for good j with the product characteristics (X_t, P_t, Ξ_t) is as given in (2.3). Since $D_t = X_t^{(1)} + \Xi_t$, the demand in market t with $(X_t^{(2)}, P_t, D_t) = (x^{(2)}, p, \delta)$ is given by:

$$\phi_j(x^{(2)}, p, \delta) = \int 1\{x_j^{(2)\prime} b^{(2)} + ap_j > -\delta_j\} 1\{(x_j^{(2)} - x_1^{(2)})' b^{(2)} + a(p_j - p_1) > -(\delta_j - \delta_1)\} \\ \cdots 1\{(x_j^{(2)} - x_J^{(2)})' b^{(2)} + a(p_j - p_J) > -(\delta_j - \delta_J)\} f_\theta(b^{(2)}, a) d\theta \ . \quad (2.17)$$

For any subset $\mathcal{J}$ of $\{1, \cdots, J\} \setminus \{j\}$, let $\mathcal{M}_\mathcal{J}$ denote the map $(x^{(2)}, p, \delta) \mapsto (\acute{x}^{(2)}, \acute{p}, \acute{\delta})$ that is uniquely defined by the following properties:

$$(\acute{x}_j^{(2)} - \acute{x}_i^{(2)}, \acute{p}_j - \acute{p}_i, \acute{\delta}_j - \acute{\delta}_i) = -(x_j^{(2)} - x_i^{(2)}, p_j - p_i, \delta_j - \delta_i), \ \forall i \in \mathcal{J} \ , \quad (2.18)$$

$$(\acute{x}_i^{(2)}, \acute{p}_i, \acute{\delta}_i) = (x_i^{(2)}, p_i, \delta_i), \ \forall i \notin \mathcal{J} \ . \quad (2.19)$$

We then define

$$\tilde{\Phi}_j(x_j^{(2)}, p_j, \delta_j) \equiv - \sum_{\mathcal{J} \subseteq \{1, \cdots J\} \setminus \{j\}} \phi_j \circ \mathcal{M}_\mathcal{J}(x^{(2)}, p, \delta) \ . \quad (2.20)$$

Eq (2.20) combines the structural demand function for good j in different markets to define a function which can be related to the random coefficient density in a simple way. This operation can be easily understood when $J = 2$, where for example ϕ_1 is given by

$$\phi_1(x^{(2)}, p, \delta) = \int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\} \\ \times 1\{(x_1^{(2)} - x_2^{(2)})' b^{(2)} + a(p_1 - p_2) < -(\delta_1 - \delta_2)\} f_\theta(b^{(2)}, a) d\theta \ . \quad (2.21)$$

Then, $\tilde{\Phi}_1$ is given by

$$\begin{aligned} \tilde{\Phi}_1(x_1^{(2)}, p_1, \delta_1) &= -\phi_1 \circ \mathcal{M}_\emptyset(x_1^{(2)} - x_2^{(2)}, p_1 - p_2, \delta_1 - \delta_2) - \phi_1 \circ \mathcal{M}_{\{2\}}(x_1^{(2)} - x_2^{(2)}, p_1 - p_2, \delta_1 - \delta_2) \\ &= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\} \Big(1\{(x_1^{(2)} - x_2^{(2)})' b^{(2)} + a(p_1 - p_2) < -(\delta_1 - \delta_2)\} \\ &\quad + 1\{(x_1^{(2)} - x_2^{(2)})' b^{(2)} + a(p_1 - p_2) > -(\delta_1 - \delta_2)\} \Big) f_\theta(b^{(2)}, a) d\theta \\ &= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\} f_\theta(b^{(2)}, a) d\theta \end{aligned} \quad (2.22)$$

This shows that aggregating the demand in the two markets with $(X_t^{(2)}, P_t, D_t) = (x^{(2)}, p, \delta)$ and $(X_{t'}^{(2)}, P_{t'}, D_{t'}) = (\acute{x}^{(2)}, \acute{p}, \acute{\delta})$ yields $\tilde{\Phi}_1$ which depends only on the utility from products 1 and 0, where the latter is normalized to 0. Eq. (2.20) generalizes this argument to more general settings with $J \geq 2$.

Under the condition we provide below, there is a product $l = j$ and a function $\tilde{\Phi}_l$, which can be written in general as $\tilde{\Phi}_l(X_j^{(2)}, P_j, D_j) = \int 1\{X_j^{(2)\prime(2)} + aP_j < -D_j\} f_\theta(b^{(2)}, a) d\theta$. Let $w \equiv (x_j^{(2)}, p_j)/\|(x_j^{(2)}, p_j)\|$ and $u \equiv \delta_j/\|(x_j^{(2)}, p_j)\|$. Define

$$\Phi_l(w, u) \equiv \tilde{\Phi}_l\Big(\frac{x_j^{(2)}}{\|(x_j^{(2)}, p_j)\|}, \frac{p_j}{\|(x_j^{(2)}, p_j)\|}, \frac{\delta_j}{\|(x_j^{(2)}, p_j)\|}\Big) = \tilde{\Phi}_l(x_j^{(2)}, p_j, \delta_j), \tag{2.23}$$

where the second equality holds because normalizing the scale of $(x_j^{(2)}, p_j, \delta_j)$ does not change the value of $\tilde{\Phi}_l$. Φ_l then satisfies

$$\begin{aligned}\Phi_l(w, u) &= -\int 1\{w'\theta < -u\} f_\theta(b^{(2)}, a) d\theta \\ &= -\int_{-\infty}^{-u} \int_{P_{w,r}} f_\theta(b^{(2)}, a) d\mu_{w,r}(b^{(2)}, a) dr = -\int_{-\infty}^{-u} \mathcal{R}[f_\theta](w, r) dr \ , \end{aligned} \tag{2.24}$$

Hence, by taking a derivative with respect to u, we may relate Φ_l to f_θ through the Radon transform:

$$\frac{\partial \Phi_l(w, u)}{\partial u} = \mathcal{R}[f_\theta](w, u). \tag{2.25}$$

Note that since the structural demand ϕ is identified by Theorem 2.2, Φ_l is nonparametrically identified as well. Hence, Eq. (2.25) gives an operator that maps the random coefficient density to an object identified by the moment condition studied in the previous section. The following condition then ensures identification of f_θ.

Assumption 2.5. *(i) For any $\mathcal{J} \subseteq \{1, \cdots, J\} \setminus \{j\}$ and any $(x^{(2)}, p, \delta) \in supp(X_t^{(2)}, P_t, D_t)$, we have $\mathcal{M}_\mathcal{J}(x^{(2)}, p, \delta) \in supp(X_t^{(2)}, P_t, D_t)$.*

Assumption 2.5 is a condition needed to ensure that Φ_l is well defined. It requires that for any$(x^{(2)}, p, \delta) \in supp(X_t^{(2)}, P_t, D_t)$ and $\mathcal{J} \subseteq \{1, \cdots, J\} \setminus \{j\}$, the operation $\mathcal{M}_\mathcal{J}$ gives another point in the support. A full support assumption on $(X_{jt}^{(2)}, P_{jt}, D_{jt})$ is sufficient for this condition.[4]

Theorem 2.3. *Suppose Assumptions 2.1-2.5 hold. Then, f_θ is identified in Example 1, where $\theta = (\beta^{(2)}, \alpha)$.*

Example 2 (Bundles, continued)**.** Identification strategies depend on whether the two goods are substitutes or complements. First, we make the following assumption and analyze the case of substitutes.

[4]Our identification argument based on Assumption 2.5 constructs Φ_l without relying on any "thin" (lower-dimensional) subset of the support of $(X_{jt}^{(2)}, P_{jt}, D_{jt})$. If Assumption 2.5 does not hold, however, one may alternatively rely on an identification argument that uses a lower dimensional subset and retain the same identification result. This alternative approach will be discussed in the next example.

Assumption 2.6. $\Delta \leq 0$ *with probability 1.*

Under this assumption, products 1 and 2 are substitutes (Gentzkow, 2007). We may then use demand on bundles (0,0) and (1,1) to identify the random coefficient density. The demand for bundle (0,0) then reduces to

$$\begin{aligned}\phi_{(0,0)}(x^{(2)}, p, \delta) = &\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\} 1\{x_2^{(2)\prime} b^{(2)} + ap_2 < -\delta_2\} \\ &\times 1\{(x_1^{(2)} + x_2^{(2)})' b^{(2)} + a(p_1 + p_2) + \Delta < -\delta_1 - \delta_2\} f_\theta(b^{(2)}, a, \Delta) d\theta \\ = &\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\} 1\{x_2^{(2)\prime} b^{(2)} + ap_2 < -\delta_2\} f_\theta(b^{(2)}, a, \Delta) d\theta \ , \quad (2.26)\end{aligned}$$

where the second equality follows from Assumption 2.6. Given product $j \in \{1, 2\}$, let $-j$ denote the other product. Let then $\mathcal{N}_{-j}$ denote the map $(x^{(2)}, p, \delta) \mapsto (\acute{x}^{(2)}, \acute{p}, \acute{\delta})$ that is uniquely defined by the following properties:

$$(\acute{x}_{-j}^{(2)}, \acute{p}_{-j}, \acute{\delta}_{-j}) = -(x_{-j}^{(2)}, p_{-j}, \delta_{-j}) \ , \quad \text{and} \quad (\acute{x}_j^{(2)}, \acute{p}_j, \acute{\delta}_j) = (x_j^{(2)}, p_j, \delta_j) \ . \quad (2.27)$$

We then define Φ_l as in (2.23) with $l = (0, 0)$, where

$$\tilde{\Phi}_{(0,0)}(x_j^{(2)}, p_j, \delta_j) \equiv -\phi_{(0,0)}(x^{(2)}, p, \delta) - \phi_{(0,0)} \circ \mathcal{N}_{-j}(x^{(2)}, p, \delta) \ . \quad (2.28)$$

Similar to the map $\mathcal{M}_{\mathcal{J}}$ in the previous section, $\mathcal{N}_{-j}$ combines markets with different characteristics to produce a function that is related to the joint density $f_{\beta^{(2)},\alpha}$ of $(\beta^{(2)}, \alpha)$ through a Radon transform.[5] Arguing as in (2.22), it is straightforward to show that $\partial \Phi_{(0,0)}(w, u)/\partial u = \mathcal{R}[f_{\beta^{(2)},\alpha}](w, u)$ with $w \equiv (x_j^{(2)}, p_j)/\|(x_j^{(2)}, p_j)\|$ and $u \equiv \delta_j/\|(x_j^{(2)}, p_j)\|$. Hence, one may identify $f_{\beta^{(2)},\alpha}$ by inverting the Radon transform under the following assumption.

Assumption 2.7. *(i) For any* $(x^{(2)}, p, \delta) \in supp(X_t^{(2)}, P_t, D_t)$*, we have* $\mathcal{N}_{-j}(x^{(2)}, p, \delta) \in supp(X_t^{(2)}, P_t, D_t)$*.*

If the researcher is only interested in the distribution of $(\beta^{(2)}, \alpha)$ but not in the bundle effect, the demand for $(0, 0)$ is enough for recovering their density. However, Δ is often of primary interest. The demand on (1,1) can be used to recover its distribution by the following argument.

Under Assumption 2.6, an argument similar to (2.26) yields

$$\begin{aligned}&\phi_{(1,1)}(x^{(2)}, p, \delta) \\ &\quad = \int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + \Delta > -\delta_1\} 1\{x_2^{(2)\prime} b^{(2)} + ap_2 + \Delta > -\delta_2\} f_\theta(b^{(2)}, a, \Delta) d\theta. \quad (2.29)\end{aligned}$$

[5]Since the bundle effect Δ does not appear in (2.26), one may only identify the joint density of the subvector $(\beta^{(2)}, \alpha)$ from the demand for bundle (0,0).

Note that Δ can be viewed as a random coefficient on the constant whose sign is fixed. Hence, the set of covariates includes a constant. In this setting, the map $\mathcal{N}_j$ cannot be used to yield a function that depends only on $(x_j^{(2)}, p_j, \delta_j)$. This is because Assumption 2.7 cannot be satisfied due to the presence of the constant term, i.e. $\mathcal{N}_{-j}(x^{(2)}, p, \delta, 1) \notin supp(X_t^{(2)}, P_t, D_t, 1)$. Hence, we consider alternative approaches.

One approach would be to assume that at least one of D_{jt}s, say D_{-jt}, has a large support so that $X_{-jt}^{(2)\prime} b^{(2)} + \alpha P_{-jt} + \Delta > -D_{-jt}$ with probability 1 provided that D_{-jt} being sufficiently large. Another alternative approach would be to condition on the set of covariates such that the observable characteristics of the two products except $X_{jt}^{(1)}$ are equal to each other.[6] That is,

$$x_1^{(2)} = x_2^{(2)} = \bar{x}^{(2)} \ , \quad \text{and} \quad p_1 = p_2 = \bar{p} \ , \tag{2.30}$$

for some $(\bar{x}^{(2)}, \bar{p})$. This then reduces (2.29) to

$$\phi_{(1,1)}(x^{(2)}, p, \delta) = \int 1\{x_1^{(2)\prime} b^{(2)} + a p_1 + \Delta > \max\{-\delta_1, -\delta_2\}\} f_\theta(b^{(2)}, a, \Delta) d\theta \ .$$

Conditioning on the markets with $\delta_1 < \delta_2$ and normalizing the arguments by the norm of $(x^{(2)}, p)$ yield a function $\Phi_{(1,1)}$ that is related to f_θ as in (2.16). Either of the two approaches uses a lower dimensional subset of the support of the covariates. Hence, care must be taken in practice to decide which approach would retain more information in the data.

For simplicity, suppose that we take the first approach and assume that D_{-jt} has a large enough support.

Theorem 2.4. *Suppose Assumption 2.1-2.4, and 2.6 hold. (a) If Assumption 2.7 holds, then $f_{\beta^{(2)},\alpha}$ is nonparametrically identified in Example 2 where the two goods are substitutes; (b) If D_{-jt} has a large enough support, then f_θ is nonparametrically identified in Example 2 where the two goods are substitutes, and $\theta = (\beta^{(2)}, \alpha, \Delta)$.*

We also consider identification of f_θ when the products are complements.

Assumption 2.8. $\Delta \geq 0$ *with probability 1.*

Under this assumption, products 1 and 2 are complements. Let $\tilde{\mathbb{L}} = \{(1,0), (0,1)\}$. The

[6] In practice, one should look for markets where the two products are similar in terms of $(X_{jt}^{(2)}, P_{jt})$.

demand for bundles (1,0) and (0,1) then reduce to

$$\phi_{(1,0)}(x^{(2)},p,\delta)=\int 1\{x_1^{(2)\prime}b^{(2)}+ap_1>-\delta_1\}1\{x_2^{(2)\prime}b^{(2)}+ap_2+\Delta<-\delta_2\}f_\theta(b^{(2)},a,\Delta)d\theta\ , \tag{2.31}$$

$$\phi_{(0,1)}(x^{(2)},p,\delta)=\int 1\{x_2^{(2)\prime}b^{(2)}+ap_2>-\delta_2\}1\{x_1^{(2)\prime}b^{(2)}+ap_1+\Delta<-\delta_1\}f_\theta(b^{(2)},a,\Delta)d\theta\ , \tag{2.32}$$

Hence, if Assumption 2.7 holds, by arguing as in the case of substitutes, we may use $\phi_{(1,0)}$ or $\phi_{(0,1)}$ to recover f_θ. For example, suppose Assumption 2.7 holds with $j=1$ and $-j=2$. Then, one may construct

$$\tilde{\Phi}_{(0,1)}(x_1^{(2)},p_1,\delta_1)=-\phi_{(0,1)}(x^{(2)},p,\delta)-\phi_{(0,1)}\circ\mathcal{N}_2(x^{(2)},p,\delta). \tag{2.33}$$

Define Φ_l as in (2.23) with $l=(0,1)$ and arguing as before, it is then straightforward to show that $\partial\Phi_{(0,1)}(w,u)/\partial u=\mathcal{R}[f_\theta](w,u)$ with $w=(x_1^{(2)},p_1,1)/\|(x_1^{(2)},p_1,1)\|$ and $u=-\delta_1/\|(x_1^{(2)},p_1,1)$. We summarize the identification result below.

Theorem 2.5. *Suppose Assumptions 2.1-2.4, 2.7, and 2.8 hold, then f_θ is nonparametrically identified in Example 2, where the two goods are complements and $\theta=(\beta^{(2)},\alpha,\Delta)$.*

Example 3 (Multiple units, continued)**.** The nonparametric IV step identifies ϕ_l for $l\in\{(0,0),(0,1),(2,0),(2,1)\}$. Using them, we may first recover the joint density of some of the random coefficients: $\theta=(\beta^{(2)},\alpha,\Delta_{(1,1)},\Delta_{(2,0)},\Delta_{(2,1)})'$. We begin with the demand for $(0,1)$, $(2,0)$, and $(2,1)$ given by

$$\begin{aligned}
\phi_{(0,1)}(x^{(2)},p,\delta)=&\int 1\{x_2^{(2)\prime}b^{(2)}+ap_2>-\delta_2\}\\
&\times 1\{(x_1^{(2)}-x_2^{(2)})'b^{(2)}+a(p_1-p_2)<-\delta_1+\delta_2\}\\
&\times 1\{x_1^{(2)\prime}b^{(2)}+ap_1+\Delta_{(1,1)}>-\delta_1\}f_\theta(b^{(2)},a,\Delta_{(1,1)},\Delta_{(2,0)},\Delta_{(2,1)})d\theta\ ,\\
\phi_{(2,0)}(x^{(2)},p,\delta)=&\int 1\{x_1^{(2)\prime}b^{(2)}+ap_1+\Delta_{(2,0)}>-\delta_1\}\\
&\times 1\{(x_1^{(2)}-x_2^{(2)})'b^{(2)}+a(p_1-p_2)+\Delta_{(2,0)}-\Delta_{(1,1)}>-\delta_1+\delta_2\}\\
&\times 1\{x_2^{(2)\prime}b^{(2)}+ap_2+\Delta_{(2,1)}-\Delta_{(2,0)}<-\delta_2\}f_\theta(b^{(2)},a,\Delta_{(1,1)},\Delta_{(2,0)},\Delta_{(2,1)})d\theta\\
\phi_{(2,1)}(x^{(2)},p,\delta)=&\int 1\{x_1^{(2)\prime}b^{(2)}+ap_1+\Delta_{(2,1)}-\Delta_{(1,1)}>-\delta_1\}\\
&\times 1\{x_1^{(2)\prime}b^{(2)}+ap_1+\Delta_{(2,1)}-\Delta_{(2,0)}>-\delta_2\}\\
&\times 1\{(x_1^{(2)}+x_2^{(2)})'b^{(2)}+a(p_1+p_2)+\Delta_{(2,1)}<-\delta_1-\delta_2\}f_\theta(b^{(2)},a,\Delta_{(1,1)},\Delta_{(2,0)},\Delta_{(2,1)})d\theta\ .
\end{aligned}$$

Hence, if D_2 has a large support, by taking δ_2 sufficiently small, we may define

$$\begin{aligned}
\tilde{\Phi}_{(0,1)}(x_1^{(2)}, p_1, \delta_1) &\equiv -\phi_{(0,1)}(x^{(2)}, p, \delta) \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + \Delta_{(1,1)} > -\delta_1\} f_\theta(b^{(2)}, a, \Delta_{(1,1)}, \Delta_{(2,0)}, \Delta_{(2,1)}) d\theta \ , \qquad (2.34) \\
\tilde{\Phi}_{(2,0)}(x_1^{(2)}, p_1, \delta_1) &\equiv -\phi_{(2,0)}(x^{(2)}, p, \delta) \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + \Delta_{(2,0)} > -\delta_1\} f_\theta(b^{(2)}, a, \Delta_{(1,1)}, \Delta_{(2,0)}, \Delta_{(2,1)}) d\theta \ , \qquad (2.35) \\
\tilde{\Phi}_{(2,1)}(x_1^{(2)}, p_1, \delta_1) &\equiv -\phi_{(2,1)}(x^{(2)}, p, \delta) \\
&= \int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + \Delta_{(2,1)} - \Delta_{(1,1)} > -\delta_1\} f_\theta(b^{(2)}, a, \Delta_{(1,1)}, \Delta_{(2,0)}, \Delta_{(2,1)}) d\theta \ . \qquad (2.36)
\end{aligned}$$

Then, defining $\Phi_l, l \in \{(0,1),(2,0),(2,1)\}$ as in (2.23) and arguing as in Example 2, Φ_l is related to the random coefficient densities by

$$\frac{\partial \Phi_l(w,u)}{\partial u} = \mathcal{R}[f_l](w,u), \quad l \in \{(0,1),(2,0),(2,1)\},$$

where $w \equiv -(x_1^{(2)}, p_1, 1)/\|(x_1^{(2)}, p_1, 1)\|$ and $u \equiv \delta_1/\|(x_1^{(2)}, p_1, 1)\|$. Here, f_ls are proper joint densities of subcomponents of θ, which are given by[7]

$$f_{(0,1)} = f_{(\beta^{(2)}, \alpha, \Delta_{(1,1)})}, \ f_{(2,0)} = f_{(\beta^{(2)}, \alpha, \Delta_{(2,0)})}, \ f_{(2,1)} = f_{(\beta^{(2)}, \alpha, \Delta_{(2,1)} - \Delta_{(1,1)})}. \qquad (2.37)$$

The joint density of θ is identified by making the following assumption.

Assumption 2.9. *(i)* $(\Delta_{(1,1)}, \Delta_{(2,0)}, \Delta_{(2,1)})$ *are independent of each other conditional on* $(\beta^{(2)}, \alpha)$*; (ii) The characteristic function of* $\Delta_{(1,1)}$ *conditional on* $(\beta^{(2)}, \alpha)$ *is nonzero a.e.*

Assumption 2.9 (i) means that, relative to the benchmark utility given as an index function of $(X_t^{(2)}, P_t, D_t)$, the additional utilities from the bundles are independent of each other. Assumption 2.9 (ii) is a regularity condition for recovering the distribution of $\Delta_{(2,1)}$ from those of $\Delta_{(2,1)} - \Delta_{(1,1)}$ and $\Delta_{(1,1)}$ through deconvolution.

Identification of the joint density f_θ allows one to recover the demand for the middle alternative: (1,0), which remained unidentified in our analysis in Section 2.2. To see this, we note

[7] Alternative assumptions can be made to identify the joint density of different components of the random coefficient vector. For example, a large support assumption on D_1 would allow one to recover the joint density of $(\beta^{(2)}, \alpha, \Delta_{(2,1)} - \Delta_{(2,0)})$ from the demand for bundle (2,0).

that the demand for this bundle is given by

$$\begin{aligned}\phi_{(1,0)}(x^{(2)},p,\delta) = \int & 1\{0 < x_1^{(2)\prime}b^{(2)} + ap_1 + \delta_1 < -\Delta_{(2,0)}\} \\ & \times 1\{x_2^{(2)\prime}b^{(2)} + ap_2 + \delta_2 < -\Delta_{(1,1)}\}1\{(x_1^{(2)} - x_2^{(2)})'b^{(2)} + a(p_1 - p_2) < -(\delta_1 - \delta_2)\} \\ & \times 1\{(x_1^{(2)} + x_2^{(2)})'b^{(2)} + a(p_1 + p_2) + \Delta_{(2,1)} < -(\delta_1 + \delta_2)\} f_\theta(b^{(2)}, a, \Delta_{(1,1)}, \Delta_{(2,0)}, \Delta_{(2,1)})d\theta.\end{aligned} \tag{2.38}$$

Since the previously unknown density f_θ is identified, this demand function is identified. This and $\phi_{(1,1)} = 1 - \sum_{l \in \mathbb{L}\setminus\{(1,1)\}} \phi_l$ further imply that all components of ϕ are now identified. We summarize these results below as a theorem.

Theorem 2.6. *Suppose $U_{(y_1,y_2),t}$ is concave in (y_1, y_2). Suppose Condition 2.1 and Assumptions 2.1, 2.3-2.4 hold. Suppose further that Assumption 2.9 holds. Then, f_θ is identified in Example 3, where $\theta = (\beta^{(2)}, \alpha, \Delta_{(1,1)}, \Delta_{(2,0)}, \Delta_{(2,1)})'$. Further, all components of the structural demand ϕ are identified.*

3 Extensions

3.1 Alternative specific coefficients

So far, we have maintained Assumption 2.1 (ii), which requires $(\beta_j, \alpha_j) = (\beta, \alpha), a.s.$ for all j. This excludes alternative specific random coefficients. However, this is not essential in our analysis. One may allow some or all components of (β_j, α_j) to be different random variables across j and identify their joint distribution under an extended support condition on the product characteristics. A leading example is that the utility contains an alternative specific additive unobserved heterogeneity term ϵ_j:

$$U_{jt}^* = X_{jt}^{(1)}\beta^{(1)} + X_{jt}^{(2)\prime}\beta^{(2)} + \alpha P_{jt} + \epsilon_j + \Xi_{jt},$$

In this setting, one has $\beta_j = (\beta^{(1)}, \beta^{(2)}, \epsilon_j)$ and $X_{jt} = (X_{jt}^{(1)}, X_{jt}^{(2)\prime}, 1)'$, and therefore, only the last component of β_j differs across alternatives. Following Berry and Pakes (2007), we call $\{\epsilon_j, j = 1, \cdots, J\}$ unobserved tastes for products.[8] Theorem 2.2 indeed still holds if Assumption 2.1 (ii) is dropped. Hence, the identification of the structural demand does not require any major modification. Another point to note is that the argument does not change even if the unobserved heterogeneity also varies across markets in an *iid* manner. That is, $(\beta_{jt}^{(2)}, \alpha_{jt})$ varies

[8]Berry and Pakes (2007) provide detailed discussions on how a model with unobserved tastes for products differs from the pure characteristic model in terms of allowed substitution patterns and welfare implications.

across individuals and markets independently and follows an identical distribution $f_{(\beta_j^{(2)},\alpha_j)}$. For example, it is commonly assumed that the additive unobserved error $\epsilon_t \equiv (\epsilon_{1t}, \cdots, \epsilon_{Jt})$ varies across individuals and markets independently following some joint distribution f_ϵ (e.g. independent Type-I extreme value distributions, typically denoted ϵ_{ijt} in the literature), but this does not change the conclusion of Theorem 2.2. This is because the identification argument in Theorem 2.2 is based on the properties of the aggregate demand ϕ. These properties, including the invertibility of ϕ, depend on the distribution of the random coefficients but not on particular realizations.

Given that the aggregate demand is identified as before, the marginal densities $f_{(\beta_j^{(2)},\alpha_j)}$ can be identified for all js that satisfy Assumption 2.4 (ii) using the same identification strategy in the previous section. For this, we note that the maps $\mathcal{M}_\mathcal{J}$ or $\mathcal{N}_{-j}$ cannot be used as the use of these maps are justified by Assumption 2.1 (ii). However, the large support assumption on $D_i, i \neq j$ can be used to construct Φ_j.

Identification of the joint density f_θ requires an extension of our identification strategy. To see this, we take Example 2 as an illustration below. Consider identifying the joint density of $\theta = (\beta_1^{(2)}, \beta_2^{(2)}, \alpha_1, \alpha_2, \Delta)$ under Assumption 2.8. In this setting, we may use the demand for $(1, 0)$, which can be written as

$$\phi_{(1,0)}(x^{(2)}, p, \delta) = \int 1\{x_1^{(2)\prime} b_1^{(2)} + a_1 p_1 > -\delta_1\} \times 1\{x_2^{(2)\prime} b_2^{(2)} + a_2 p_2 + \Delta < -\delta_2\} f_\theta(b_1^{(2)}, b_2^{(2)}, a_1, a_2, \Delta) d\theta. \quad (3.1)$$

To recover the joint density, one has to directly work with this demand function without simplifying it further. A key feature of (3.1) is that it involves multiple indicator functions and that distinct subsets of θ show up in each of these indicator functions. For example, the first indicator function in (3.1) involves $(\beta_1^{(2)}, \alpha_1)$, while the second indicator function involves $(\beta_2^{(2)}, \alpha_2, \Delta)$. Integral transforms of this form are studied in Dunker, Hoderlein, and Kaido (2013) in their analysis of random coefficients discrete game models. They use tensor products of integral transforms to study nonparametric identification of random coefficient densities. Using their framework, one may show that

$$\frac{\partial^2 \phi_{(1,0)}(w_1, w_2, \delta_1, \delta_2)}{\partial \delta_1 \partial \delta_2} = (\mathcal{R} \otimes \mathcal{R})[f_\theta](w_1, w_2, \delta_1, -\delta_2), \quad (3.2)$$

where $w_1 = -(x_1^{(2)}, p_1)/\|(x_1^{(2)}, p_1)\|$, $w_2 = (x_2^{(2)}, p_2)/\|(x_2^{(2)}, p_2)\|$, and $\mathcal{R} \otimes \mathcal{R}$ is the tensor product of Radon transforms, which can be inverted to identify f_θ. The main principle of our identification strategy is therefore the same as before. Inverting the transform in (3.2) to identify f_θ requires Assumption 2.4 (ii) to be strengthened as follows.

Assumption 3.1. $(X_{1t}^{(2)}, P_{1t}, D_{1t}, X_{2t}^{(2)}, P_{2t}, D_{2t})$ *has a full support.*

This is a stronger support condition than Assumption 2.4 (ii) as it not only requires a full support condition for the characteristics of one good but also for the characteristics of both goods. This condition is violated, for example, when there is a common covariate that enters the characteristics of both goods. This is in line with the previous findings in the literature that identifying the joint distribution of potentially correlated unobservable tastes for products (e.g. ϵ_1 and ϵ_2) requires variables that are excluded from one or more goods (see e.g. Keane, 1992 and Gentzkow, 2007). Identification of f_θ is then established by the following theorem.[9]

Theorem 3.1. *In Example 2, let $\theta = (\beta_1, \beta_2, \alpha_1, \alpha_2, \Delta)$. Suppose that Assumptions 2.1 (i), 2.2-2.3 hold. Then, ψ is identified. Suppose further that Assumptions 2.4 (i), (iii), 2.8 and 3.1 hold. Then, f_θ is identified.*

3.2 Nonparametric identification of ψ with full independence

In Section 2.2, we discussed the the nonparametric identification of the functions ψ_j in the equation $\Xi_{jt} = \psi_j(X_t^{(2)}, P_t, \tilde{S}_t) - X_{jt}^{(1)}$. Following BH (2013), we proposed to identify the structural functions by the conditional moment equations

$$E\Big[\psi_j\left(X_t^{(2)}, P_t, S_t\right)\Big| Z_t = z_t, X_t = \left(x_t^{(1)}, x_t^{(2)}\right)\Big] = x_{jt}^{(1)}, \quad j = 1, \cdots, J.$$

with instrumental variables Z_t. The identification relies on the assumption that the unobservable Ξ_{jt} is mean independent of the instruments. However, in many applications researchers choose instruments by arguing that they are independent of the unobservable. Using only mean independence means using only parts of the available information. Thereby, the identifying power is weakened. Adding the stronger independence assumption when it is justified will improve identification as well as estimation. Therefore, we propose an approach similar to Dunker et. al. (2014) by formally assuming

$$\Xi_{jt} \perp\!\!\!\perp (Z_t, X_t) \quad \text{and } E[\Xi_{jt}] = 0 \qquad \text{for all } j, t.$$

This leads to the nonlinear equation

$$0 = \begin{pmatrix} P[\psi_j(X_t^{(2)}, S_t, P_t) - X_{jt}^{(1)} \leq \xi] - P[\psi_j(X_t^{(2)}, S_t, P_t) - X_{jt}^{(1)} \leq \xi | Z_t = z_t, X_t = x_t] \\ E[\psi_j(X_t^{(2)}, S_t, P_t) - X_{jt}^{(1)}] \end{pmatrix}$$

for all ξ, z_t, x_t. Nonparametric estimation of problems involving this type of nonlinear restrictions are studied in Chen and Pouzo (2012) and Dunker et. al. (2014). To give sufficient

[9] We omit the proof of this result for brevity.

conditions for identification, we define the operator

$$F(\varphi)(\xi, z_t, x_t) := \begin{pmatrix} P[\varphi(X_t^{(2)}, S_t, P_t) - X_{jt}^{(1)} \leq \xi] - P[\varphi(X_t^{(2)}, S_t, P_t) - X_{jt}^{(1)} \leq \xi | Z_t = z_t, X_t = x_t] \\ E[\varphi(X_t^{(2)}, S_t, P_t) - X_{jt}^{(1)}] \end{pmatrix}.$$

The function ψ_j is a root of the operator F. It is, therefore, globally identified under the following assumption.

Assumption 3.2. *The operator F has a unique root.*

On first sight this may appear as a strong assumption due to the complexity of the operator. It is, however, weaker than the usual completeness assumption for the mean independence assumption. This is because, if $\Xi_{jt} \perp\!\!\!\perp (Z_t, X_t)$ and the usual completeness assumption hold, then F has only one root. On the other hand, completeness is not necessary for F to have a unique root. Hence, when $\Xi_{jt} \perp\!\!\!\perp (Z_t, X_t)$, Assumption 3.2 is weaker than Assumption 2.3. Another important advantage of this method is that because the D_j do not vanish, we have a close analog to nonparametric IV with full independence, see, e.g., Dunker et al (2014), where D_j now plays the role of the dependent variable.

3.3 Covariates with bounded support

Our basic identification strategy is to invert the Radon transform in (2.15). This inversion requires rich variation in w. In practice, however, one may not be able to vary w flexibly when $(X_j^{(2)}, P_j)$ has a limited support. Even in such a setting, identification of f_θ is still possible under an additional assumption on the moments of the random coefficients. To cover all examples considered, we let θ_j denote the j-th component of θ.

Assumption 3.3. *All the absolute moments of $\theta_j, j = 1, \cdots, d_\theta$ are finite, and for any fixed $z \in \mathbb{R}_+$, $\lim_{l\to\infty} \frac{z^l}{l!}(E[|\theta_1|^l] + \cdots + E[|\theta_{d_\theta}|^l]) = 0$.*

Under Assumption 3.3, the characteristic function $w \mapsto \varphi_\theta(tw)$ of θ (a key element of the Radon inversion) is uniquely determined by its restriction to a non-empty full dimensional subset of $\mathbb{S}^{d_\theta}$.[10] Hence, f_θ can still be identified if one may vary w on a non-empty full dimensional subset. For example, if $(X_j^{(2)}, P_j)$'s support contains an open ball in $\mathbb{R}^{k_2+1}$, this is sufficient for the identification of f_θ.

[10] This follows from analytic continuation. See Hoderlein, Holzman, and Meister (2014) and Masten (2014) for details.

4 Suggested estimation methods

4.1 Nonparametric estimator

The structure of the nonparametric identification suggests a nonparametric estimation strategy in a natural way. It consists of three steps. The first step is the estimation of the structural function ψ_j. The second step is to derive the function Φ_j from the estimated $\widehat{\psi}_j$. This requires only straightforward algebraic manipulation which were presented in Section 2.3. We will not further comment on these computations. The last step of the estimation is the inversion of a Radon transform.

The mathematical structure of the first step is similar to nonparametric IV. The conditional expectation operator on the left hand side of the equation

$$E[\psi_j(x_t^{(2)}, P_t, S_t)|Z_t = z_t, X_t = x_t] = x_{jt}^{(1)} \qquad \text{for all } x_t, z_t$$

has to be inverted. Let us denote this linear operator by T and rewrite the problem as $(T\psi_j)(z_t, x_t) = x_{jt}^{(1)}$. Here $x_{jt}^{(1)}$ should be interpreted as a function in x_t and z_t which is constant in $x_t^{(2)}$, z_t, and $x_{it}^{(1)}$ for $i \neq j$. The operator depends on the joint density of (X_t, P_t, S_t, Z_t) which has to be estimated nonparametrically, e.g. by kernel density estimation. This gives an estimator $\widehat{T}$. As in nonparametric IV the operator equation is usually ill-posed. Regularized inversion schemes must be applied. We propose Tikhonov regularization for this purpose:

$$\widehat{\psi}_j := \min_{\psi} \|\widehat{T}\psi - x_{jt}^{(1)}\|^2_{L^2(X_t,Z_t)} + \alpha\mathfrak{R}(\psi). \tag{4.1}$$

As usual $\alpha \geq 0$ is a regularization parameter and $\mathfrak{R}$ a regularization functional. The usual choice would be $\mathfrak{R}(\psi) = \|\psi\|^2_{L^2}$. If more smoothness is expected, this could be a squared Sobolev norm or some other norm as well. In the case of bundles and multiple goods we know that ψ must be monotonically increasing or decreasing in S_t. One may incorporate this a priori knowledge by setting $\mathfrak{R}(\psi) = \infty$ for all functions ψ not having this property. Monotonicity is a convex constraint. Hence, even with this choice of $\mathfrak{R}$, equation (4.1) is a convex minimization problem. Solving the problem is computationally feasible. We refer to Eggermont (1993), Burger and Osher (2004), and Resmerita (2005) for regularization with general convex regularization functional. Furthermore, we refer to Newey and Powell (2003) for the related nonparametric IV problem.

The third step of our nonparametric estimation strategy is the inversion of a Radon transform. A popular and efficient method for the problem is the filtered back projection

$$\widehat{f}_\theta(\vartheta) = \mathcal{R}^* \left(\Omega_r *_\delta \frac{\partial \Phi_j(x_j^{(2)}, p_j, \delta_j)}{\partial \delta_j} \right)(\vartheta).$$

Here $\vartheta = (b, a)$ or (b, a, Δ) depending on the application, $(R^*g)(x) := \int_{\|w\|=1} g(w, w'x)dw$ is the adjoint of the Radon transform, and $*_\delta$ denotes the convolution with respect to the last variable δ_j, and Ω_r is the function

$$\Omega_r(s) := \frac{1}{4\pi^2} \begin{cases} (\cos(rs) - 1)/s^2 & \text{for } s \neq 0, \\ r^2/2 & \text{for } s = 0. \end{cases}$$

For more details on this algorithm in a deterministic setting we refer to Natterer (2001). A similar estimator for random coefficients is proposed and analyzed in HKM.

4.2 Parametric estimators for bundle choice models

Our nonparametric identification analysis shows that the choice of bundles and multiple units of consumption can be studied very much in the same way as the standard BLP model (or the pure characteristic model). This suggests that one may construct parametric estimators for these models by extending standard estimation methods, given appropriate data. Below, we take Example 2 and illustrate this idea.

Let $\theta = (\beta^{(2)}, \alpha, \Delta)$ be random coefficients and let $f_\theta(\cdot; \gamma)$ be a parametric density function, where γ belongs to a finite dimensional parameter space $\Gamma \subset \mathbb{R}^{d_\gamma}$. The estimation procedure consists of the following steps:

Step 1 : Compute the aggregate share of bundles as a function of parameter γ conditional on the set of covariates.

Step 2 : Use numerical methods to solve demand systems for (D_1, D_2), where $D_j = \Xi_j + X_j^{(1)}, j = 1, 2$ and obtain the inversion in eq. (2.14).

Step 3 : Form a GMM criterion function using instruments and minimize it with respect to γ over the parameter space.

The first step is to compute the aggregate share. In the pure characteristic model, one may approximate the aggregate share of each bundle such as the one in (2.7) by simulating θ from $f_\theta(\cdot; \gamma)$ for each γ. Specifically, when the conditional CDF of α given $(\beta^{(2)}, \Delta)$ has an analytic form, the two-step method in BLP and Berry and Pakes (2007) can be employed. We take the demand for bundle (1,0) in eq. (2.7) as an example. Conditional on the product characteristics $y \equiv (x^{(2)}, p, \delta)$ and the rest of the random coefficients $(\beta^{(2)}, \Delta)$, bundle (1,0) is chosen when

$$\underline{A}(y, \beta^{(2)}, \Delta) < \alpha < \overline{A}(y, \beta^{(2)}, \Delta), \tag{4.2}$$

where

$$\underline{A}(y,\beta^{(2)},\Delta) \equiv \frac{-x_2^{(2)\prime(2)} - \Delta - \delta_2}{p_2},$$
$$\overline{A}(y,\beta^{(2)},\Delta) \equiv \min\{\frac{-x_2^{(2)\prime(2)} - \delta_2}{p_1}, \frac{(x_2^{(2)} - x_1^{(2)})'^{(2)} + (\delta_2 - \delta_1)}{p_1 - p_2}\}.$$

Let $F_\alpha(\cdot|\beta^{(2)},\Delta)$ denote the conditional CDF of α. Then, (4.2) implies that the aggregate share of bundle (1,0) is given by

$$\phi_{(1,0)}(x^{(2)},p,\delta;\gamma) = \int \left(F_\alpha(\overline{A}(y,b^{(2)},\Delta)|b^{(2)},\Delta) - F_\alpha(\underline{A}(y,b^{(2)},\Delta)|b^{(2)},\Delta)\right) \times 1\{\overline{A}(y,\beta^{(2)},\Delta) > \underline{A}(y,\beta^{(2)},\Delta)\} f_{\beta^{(2)},\Delta}(b,\Delta;\gamma)d\theta. \quad (4.3)$$

This can be approximated by the simulated moment:

$$\hat{\phi}_{(1,0)}(x^{(2)},p,\delta;\gamma) = \frac{1}{n_S}\sum_{i=1}^{n_S}\left(F_\alpha(\overline{A}(y,b_i^{(2)},\Delta_i)|b_i^{(2)},\Delta_i) - F_\alpha(\underline{A}(y,b_i^{(2)},\Delta_i)|b_i^{(2)},\Delta_i)\right) \times 1\{\overline{A}(y,b_i^{(2)},\Delta_i) > \underline{A}(y,b_i^{(2)},\Delta_i)\}, \quad (4.4)$$

where the simulated sample $\{(b_i^{(2)},\Delta_i), i = 1,\cdots,n_S\}$ is generated from $f_{\beta^{(2)},\Delta}(\cdot;\gamma)$.[11] Computation of the aggregate demand for other bundles is similar. This step therefore gives the model predicted aggregate demand $\hat{\phi}_l$ for all bundles under a chosen parameter value γ.

The next step is then to invert subsystems of demand and obtain ψ numerically. Given $\hat{\phi}_l, l \in \mathbb{L}$ from Step 1, this step can be carried out by numerically calculating inverse mappings. For example, take $\tilde{\mathbb{L}} = \{(1,0),(1,1)\}$. Then, $(\delta_1,\delta_2) \mapsto (\hat{\phi}_{(1,0)}(x^{(2)},p,\delta;\gamma), \hat{\phi}_{(1,1)}(x^{(2)},p,\delta;\gamma))$ defines a mapping from $\mathbb{R}^2$ to $[0,1]^2$. Standard numerical methods such as the Newton-Raphson method or the homotopy method (see Berry and Pakes, 2007) can then be employed to calculate the inverse of this mapping[12], which then yields $\hat{\psi}(\cdot;\gamma) \equiv (\hat{\psi}_1(\cdot;\gamma), \hat{\psi}_2(\cdot;\gamma))$ such that

$$\Xi_{1,t} = \hat{\psi}_1(X_t^{(2)},P_t,S_{(1,0),t},S_{(1,1),t};\gamma) - X_{1t}^{(1)}, \quad \Xi_{2,t} = \hat{\psi}_2(X_t^{(2)},P_t,S_{(1,0),t},S_{(1,1),t};\gamma) - X_{2t}^{(1)} \quad (4.5)$$

where $(S_{(1,0),t}, S_{(1,1),t})$ are observed shares of bundles. One may further repeat this step with $\tilde{\mathbb{L}} = \{(0,0),(0,1)\}$, which yields

$$\Xi_{1,t} = \hat{\psi}_3(X_t^{(2)},P_t,S_{(0,0),t},S_{(0,1),t};\gamma) - X_{1t}^{(1)}, \quad \Xi_{2,t} = \hat{\psi}_4(X_t^{(2)},P_t,S_{(0,0),t},S_{(0,1),t};\gamma) - X_{2t}^{(1)} \quad (4.6)$$

This helps generate additional moment restrictions in the next step.

[11] One may also use an importance sampling method.

[12] Whether the demand subsystems admit an analog of BLP's contraction mapping method is an interesting open question, which we leave for future research.

The third step is to use (4.5)-(4.6) to generate moment conditions and estimate γ by GMM. There are four equations in total, while because the shares sum up to 1 one equation is redundant. Hence, by multiplying instruments to the residuals from the first three equations, we define the sample moment:

$$g_n(X_t, P_t, S_t, Z_t; \gamma) \equiv \frac{1}{n} \sum_{t=1}^{n} \begin{pmatrix} \hat{\psi}_1(X_t^{(2)}, P_t, S_{(1,0),t}, S_{(1,1),t}; \gamma) - X_{1t}^{(1)} \\ \hat{\psi}_2(X_t^{(2)}, P_t, S_{(1,0),t}, S_{(1,1),t}; \gamma) - X_{2t}^{(1)} \\ \hat{\psi}_3(X_t^{(2)}, P_t, S_{(0,0),t}, S_{(0,1),t}; \gamma) - X_{1t}^{(1)} \end{pmatrix} \otimes \begin{pmatrix} Z_t \\ X_t \end{pmatrix}.$$

Letting $W_n(\gamma)$ be a (possibly data dependent) positive definite matrix, define the GMM criterion function by

$$Q_n(\gamma) \equiv g_n(X_t, P_t, S_t, Z_t; \gamma)' W_n(\gamma) g_n(X_t, P_t, S_t, Z_t; \gamma).$$

The GMM estimator $\hat{\gamma}$ of γ can then be computed by minimizing Q_n over the parameter space. A key feature of this method is that it uses the familiar BLP methodology (simulation, inversion & GMM) but yet allows one to estimate models that do not fall in the class of multinomial choice models. Employing our procedure may, for example, allow one to estimate bundle choices (e.g. print newspaper, online newspaper, or both) or platform choices using market level data.

5 Outlook

This paper is concerned with the nonparametric identification of models of market demand. It provides a general framework that nests several important models, including the workhorse BLP model, and provides conditions under which these models are point identified. Important conclusions include that the assumption necessary to recover various objects differ; in particular, it is easier to identify demand elasticities and more difficult to identify the individual specific random coefficient densities. Moreover, the data requirements are also shown to vary with the model considered. The identification analysis is constructive, extends the classical nonparametric BLP identification as analyzed in BH to other models, and opens up the way for future research on sample counterpart estimation. A particularly intriguing part hereby is the estimation of the demand elasticities, as the moment condition is different from the one used in nonparametric IV. Understanding the properties of these estimators, and evaluating their usefulness in an application, is an open research question that we hope this paper stimulates.

References

[1] Ackerberg, D., C. L. Benkard, S. Berry, and A. Pakes (2007): "Chapter 63 Econometric

Tools for Analyzing Market Outcomes," vol. 6, Part A of *Handbook of Econometrics*, pp. 4171 – 4276. Elsevier.

[2] Andrews, D. (2011): "Examples of L2-Complete and Boundedly-complete Distributions," Cowles Foundation Discussion Paper.

[3] Armstrong, T. (2013): "Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply," Discussion Paper, Yale University.

[4] Benning, M., and M. Burger (2011): "Error estimates for general fidelities," *Electron. Trans. Numer. Anal.*, 38, 44–68.

[5] Beran, R., and P. Hall (1992): "Estimating coefficient distributions in random coefficient regressions," *The annals of Statistics*, 20(4), 1970–1984.

[6] Berry, S. T., A. Gandhi, and P. A. Haile (2013): "Connected Substitutes and Invertibility of Demand," *Econometrica*, 81(5), 2087–2111.

[7] Berry, S. T., and P. A. Haile (2011): "Nonparametric Identification of Multinomial Choice Demand Models With Heterogeneous Consumers," Cowles Foundation Discussion Paper.

[8] ——— (2013): "Identification in Differentiated Products Markets Using Market Level Data," Cowles Foundation Discussion Paper.

[9] Berry, S. T., J. A. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 841–890.

[10] ——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112(1), 68–105.

[11] Berry, S. T., and A. Pakes (2007): "The Pure Characteristics Demand Model," *International Economic Review*, 48(4), 1193–1225.

[12] Burger, M., and S. Osher (2004): "Convergence rates of convex variational regularization," *Inverse problems*, 20(5), 1411.

[13] Carrasco, M., and J. Florens (2010): "A spectral method for deconvolving a density," *Econometric Theory*, 27(3), 546–581.

[14] Chen, X., and D. Pouzo (2012): "Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals," *Econometrica*, 80(1), 277–321.

[15] Dubé, J.-P., J. T. Fox, and C.-L. Su (2012): "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, 80(5), 2231–2267.

[16] Dunker, F., J.-P. Florens, T. Hohage, J. Johannes, and E. Mammen (2014): "Iterative Estimation of Solutions to Noisy Nonlinear Operator Equations in Nonparametric Instrumental Regression," *Journal of Econometrics*, 178(3), 444 – 455.

[17] Dunker, F., S. Hoderlein, and H. Kaido (2013): "Random Coefficients in Static Games of Complete Information," CEMMAP Working Paper.

[18] Eggermont, P. P. B. (1993): "Maximum entropy regularization for Fredholm integral equations of the first kind," *SIAM J. Math. Anal.*, 24, 1557–1576.

[19] Fox, J., and N. Lazzati (2013): "Identification of Discrete Choice Models for Bundles and Binary Games," CEMMAP Working Paper.

[20] Fox, J. T., and A. Gandhi (2012): "Using selection decisions to identify the joint distribution of outcomes," .

[21] Gale, D., and J. Nikaido (1965): "The Jacobian matrix and global univalence of mappings," *Math. Ann.*, 159(2), 81–93.

[22] Gautier, E., and S. Hoderlein (2013): "Estimating the Distribution of Treatment Effects," *Arxiv preprint arXiv:1109.0362*.

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

[24] Gentzkow, M. A. (2007): "Valuing New Goods in a Model with Complementarity: Online Newspapers," *American Economic Review*, 97(3), 713–744.

[25] Helgason, S. (1999): *The Radon Transform*. Birkhauser, Boston-Basel-Berlin, 2nd edition edn.

[26] Hoderlein, S., H. Holzmann, and A. Meister (2014): "The Triangular Model with Random Coefficients," Wroking Paper., Boston College.

[27] Hoderlein, S., J. Klemelä, and E. Mammen (2010): "Analyzing the Random Coefficient Model Nonparametrically," *Econometric Theory*, 26(03), 804–837.

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

[29] Keane, M. P. (1992): "A note on identification in the multinomial probit model," *Journal of Business & Economic Statistics*, 10(2), 193–200.

[30] Krantz, S. G., and H. R. Parks (2002): *The implicit function theorem: history, theory, and applications*. Birkhäuser, Boston, USA.

[31] Masten, M. A. (2014): "Random Coefficients on Endogenous Variables in Simultaneous Equations Models," .

[32] Matzkin, R. L. (2012): "Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity," *Journal of Econometrics*, 166(1), 106–115.

[33] McFadden, D. L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*, pp. 105–142.

[34] McFadden, D. L. (1981): "Structural Discrete Probability Models Derived from Theories of Choice," In Charles Manski and Daniel McFadden, editors, *Structural Analysis of Discrete Data and Econometric Applications* , The MIT Press, Cambridge.

[35] Moon, H. R., M. Shum, and M. Weidner (2014): "Estimation of Random Coefficients Logit Demand Models with Interactive Fixed Effects," CEMMAP Working Paper.

[36] Natterer, F. (2001): *The mathematics of computerized tomography*, vol. 32 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, Reprint of the 1986 original.

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

[38] Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

[39] Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4), 705–729.

[40] Resmerita, E. (2005): "Regularization of ill-posed problems in Banach spaces: convergence rates," *Inverse Problems*, 21(4), 1303–1314.

Appendix A: Notation and Definitions

The following is a list of notations and definitions used throughout the appendix.

$\mathbb{S}^{d_\theta-1}$: The unit sphere $\mathbb{S}^{d_\theta-1} \equiv \{v \in \mathbb{R}^{d_\theta} : \|v\| = 1\}$.
$\mathbb{H}_-$: The hemisphere $\mathbb{H}_- \equiv \{v = (v_1, v_2, \cdots, v_{d_\theta}) \in \mathbb{S}^{d_\theta-1} : v_{d_\theta} \leq 0\}$.
$P_{w,r}$: The hyperplane: $P_{w,r} \equiv \{v \in \mathbb{R}^{d_\theta} : v'w = r\}$.
$\mu_{w,r}$: Lebesgue measure on $P_{w,r}$.
$\mathcal{R}$: Radon transform: $\mathcal{R}[f](w,u) = \int_{P_{w,u}} f(v) d\mu_{w,u}(v)$.
$\mathcal{M}_{\mathcal{J}}$: The map $(x^{(2)}, p, \delta) \mapsto (\acute{x}^{(2)}, \acute{p}, \acute{\delta})$ defined for any $\mathcal{J} \subseteq \{1, \cdots, J\} \setminus \{j\}$ by
$(\acute{x}^{(2)}_j - \acute{x}^{(2)}_i, \acute{p}_j - \acute{p}_i, \acute{\delta}_j - \acute{\delta}_i) = -(x^{(2)}_j - x^{(2)}_i, p_j - p_i, \delta_j - \delta_i),\ \forall i \in \mathcal{J}$,
$(\acute{x}^{(2)}_i, \acute{p}_i, \acute{\delta}_i) = (x^{(2)}_i, p_i, \delta_i),\ \forall i \notin \mathcal{J}$.
$\mathcal{N}_{-j}$: The map $(x^{(2)}, p, \delta) \mapsto (\acute{x}^{(2)}, \acute{p}, \acute{\delta})$ defined by
$(\acute{x}^{(2)}_{-j}, \acute{p}_{-j}, \acute{\delta}_{-j}) = -(x^{(2)}_{-j}, p_{-j}, \delta_{-j})$, $(\acute{x}^{(2)}_j, \acute{p}_j, \acute{\delta}_j) = (x^{(2)}_j, p_j, \delta_j)$.

Appendix B: Proofs

Lemma 1. *Suppose the Assumptions 2.1 and Condition 2.1 hold and that ϕ_l is given as in Example 2 or Example 3 with $l \in \tilde{\mathbb{L}} = \{(0,1),(0,0)\}$. Then for all $(x^{(2)}, p) = \left(x^{(2)}_1, x^{(2)}_2, p_1, p_2\right) \in \mathbb{R}^{2k}$ with $(x^{(2)}_1, p_1) \neq (x^{(2)}_2, p_2)$ the function $\phi : \mathbb{R}^{2k} \times \mathbb{R}^2 \to [0,1]^2$ defined as*

$$\phi(x^{(2)}_1, x^{(2)}_2, p_1, p_2, d_1, d_2) \equiv \left[\phi_{(0,0)}\left(x^{(2)}_1, x^{(2)}_2, p_1, p_2, d_1, d_2\right), \phi_{(0,1)}\left(x^{(2)}_1, x^{(2)}_2, p_1, p_2, d_1, d_2\right)\right]$$

is invertible in (d_1, d_2) on any bounded subset of $\mathbb{R}^2$. This holds for other appropriate choices of $\tilde{\mathbb{L}}$ as well.

Proof of Lemma 1. We start with the observation that $\phi_{(0,0)}(x^{(2)}, p, d)$ is monotonically decreasing in d_1 and also in d_2 while $\phi_{(0,1)}(x^{(2)}, p, d)$ is monotonically decreasing in d_1 and monotonically increasing in d_2 by definition. It is straight forward to check in the definition of ϕ_l that the intersection of the sets characterized by the index functions in the integrals are non empty. Furthermore, the set depending only on d_1 and the set only depending on d_2 never completely contain one another unless the assumption $(x^{(2)}_1, p_1) \neq (x^{(2)}_2, p_2)$ is violated. Therefore, the full support of (β, α) implies that $\phi_{(0,0)}$ and $\phi_{(0,1)}$ are strictly increasing or decreasing in d_1 and d_2

$$\frac{\partial \phi_{(0,0)}(x^{(2)}, p, d)}{\partial d_1} < 0, \quad \frac{\partial \phi_{(0,0)}(x^{(2)}, p, d)}{\partial d_2} < 0, \quad \frac{\partial \phi_{(0,1)}(x^{(2)}, p, d)}{\partial d_1} < 0, \quad \frac{\partial \phi_{(0,1)}(x^{(2)}, p, d)}{\partial d_2} > 0.$$

Hence, the determinant of the Jacobian of $d \mapsto \phi(x^{(2)}, p, d)$ as well as their principle minors are strictly negative for all $d \in \text{supp}\,(D)$

$$\det(J_\phi)(x, d) = \frac{\partial \phi_{(0,0)}(x^{(2)}, p, d)}{\partial d_1} \frac{\partial \phi_{(0,1)}(x^{(2)}, p, d)}{\partial d_2} - \frac{\partial \phi_{(0,1)}(x^{(2)}, p, d)}{\partial d_1} \frac{\partial \phi_{(0,0)}(x^{(2)}, p, d)}{\partial d_2} < 0.$$

Thus, on every rectangular domain in $\mathbb{R}^2$ the assumptions of the Gale-Nikaido theorem are fulfilled. Since any bounded subset in $\mathbb{R}^2$ is contained in some rectangular domain, ϕ is invertible on any bounded subset of $\mathbb{R}^2$. ∎

Proof of Theorem 2.2. The proof of the theorem is immediate from Theorem 1 in BH (2013). We therefore give a brief sketch. By Assumptions 2.1 and 2.2, we note that there exists a function $\psi : \mathbb{R}^{Jk_2} \times \mathbb{R}^J \times \mathbb{R}^J \to \mathbb{R}^J$ such that for some subvector $\tilde{S}_t$ of S_t,

$$\Xi_{jt} = \psi_j(X_t^{(2)}, P_t, \tilde{S}_t) - X_{jt}^{(1)} \ , \ j = 1, \cdots, J,$$

and by Assumption 2.3, the following moment condition holds:

$$E[\psi_j(X_t^{(2)}, P_t, \tilde{S}_t) - X_{jt}^{(1)} | Z_t, X_t] = 0 \ .$$

Identification of ψ then follows from applying the completeness argument in the proof of Theorem 1 in BH (2013). ∎

Proof of Theorem 2.3. First, under the linear random coefficient specification, the connected substitutes assumption in Berry, Gandhi, and Haile (2013) is satisfied. By Theorem 1 in Berry, Gandhi, and Haile (2013), Assumption 2.2 is satisfied. Then, by Assumptions 2.1-2.3 and Theorem 2.2, ψ is identified. Further, the aggregate demand ϕ is identified by (2.13) and the identity $\phi_0 = 1 - \sum_{j=1}^J \phi_j$. By Assumptions 2.4-2.5, for each $(x^{(2)}, p, \delta)$ and $\mathcal{J} \subseteq \{1, \cdots, J\} \setminus \{j\}$, $\mathcal{M}_{\mathcal{J}}(x^{(2)}, p, \delta)$ is in the support of $(X^{(2)}, P_t, D_t)$. Hence, one may construct

$$\tilde{\Phi}_j(x_j^{(2)}, p_j, \delta_j) = \sum_{\mathcal{J} \subseteq \{1, \cdots, J\} \setminus \{j\}} \phi_j \circ \mathcal{M}_{\mathcal{J}}(x^{(2)}, p, \delta) = \int 1\{x_j^{(2)\prime} b^{(2)} + a p_j < -\delta_j\} f_\theta(b^{(2)}, a) d\theta, \quad (5.1)$$

where the second equality follows because of the following. First, $\mathcal{M}_{\mathcal{J}}$ replaces the indicators in ϕ_j of the form $1\{(x_j^{(2)} - x_i^{(2)})' b^{(2)} + a(p_j - p_i) < -(\delta_j - \delta_i)\}$ with $1\{(x_j^{(2)} - x_i^{(2)})' b^{(2)} + a(p_j - p_i) > -(\delta_j - \delta_i)\}$ for $i \in \mathcal{J}$. The random coefficients are assumed to be continuously distributed. We therefore have

$$1\{(x_j^{(2)} - x_i^{(2)})' b^{(2)} + a(p_j - p_i) < -(\delta_j - \delta_i)\} + 1\{(x_j^{(2)} - x_i^{(2)})' b^{(2)} + a(p_j - p_i) > -(\delta_j - \delta_i)\} = 1, \ a.s.$$

Therefore, $\sum_{\mathcal{J} \subseteq \{1, \cdots, J\}} \phi_j \circ \mathcal{M}_{\mathcal{J}}(x^{(2)}, p, \delta) = 1$. Since $\tilde{\Phi}_j$ is constructed by summing $\phi_j \circ \mathcal{M}_{\mathcal{J}}$ over subsets of $\{1, \cdots, J\}$ except $\{j\}$, we are left with the integral of the single indicator function $1\{x_j^{(2)\prime} b^{(2)} + a p_j < -\delta_j\}$ with respect to f_θ. This ensures (5.1). Now for $l = j$, define Φ_l as in (2.23). Then, it follows that

$$\begin{aligned} \Phi_l(w, u) &= -\int 1\{w'\theta < -u\} f_\theta(b^{(2)}, a) d\theta \\ &= -\int_{-\infty}^{-u} \int_{P_{w,r}} f_\theta(b^{(2)}, a) d\mu_{w,r}(b^{(2)}, a) dr = -\int_{-\infty}^{-u} \mathcal{R}[f_\theta](w, r) dr \ . \quad (5.2) \end{aligned}$$

Taking a derivative with respect to u then yields (2.25). Note that by Assumption 2.4 (i)-(ii),

$(X_j^{(2)}, P_j, D_j)$ is supported on $\mathbb{R}^{k_2} \times \mathbb{R} \times \mathbb{R}$. This implies that $\partial\Phi_l(w,u)/\partial u$ is well-defined for all $(w,u) \in \mathbb{S}^{d_\theta - 1} \times \mathbb{R}$. The conclusion of the theorem then follows from Assumption 2.4 (iii), injectivity of the Radon transform on $\mathcal{S}(\mathbb{R}^{d_\theta})$ (Theorem 2.4 in Helgason, 1999) and Φ_l being identified. ∎

Proof of Theorem 2.4. First, let $\tilde{\mathbb{L}} = \{(1,0),(1,1)\}$. By Assumptions 2.1-2.3 and Theorem 2.2, ψ is identified. Further, the aggregate demand $\{\phi_l, l = (1,0),(1,1)\}$ is identified by (2.14). Second, take $\tilde{\mathbb{L}} = \{(0,0),(0,1)\}$. Then by the same argument, the aggregate demand $\{\phi_l, l = (0,0),(0,1)\}$ is identified as well. Hence, the entire aggregate demand vector ϕ is identified.

By Assumption 2.6 $\Delta \leq 0$ almost surely. Then, the demand for bundle (0,0) satisfies (2.26). By Assumption 2.7, it then follows that

$$\begin{aligned}
&\tilde{\Phi}_{(0,0)}(x_j^{(2)}, p_j, \delta_j) = -\phi_{(0,0)}(x^{(2)}, p, \delta) - \phi_{(0,0)} \circ \mathcal{N}_{-j}(x^{(2)}, p, \delta) \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\}\big(1\{x_2^{(2)\prime} b^{(2)} + ap_2 < -\delta_2\} + 1\{x_2^{(2)\prime} b^{(2)} + ap_2 > -\delta_2\}\big) f_\theta(b^{(2)}, a, \Delta) d\theta \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\} f_{\beta^{(2)},\alpha}(b^{(2)}, a) d\theta, \quad (5.3)
\end{aligned}$$

where $\tilde{\Phi}_{(0,0)}$ is identified because it is constructed from $\phi_{(0,0)}$. The rest of the proof is then the same as the proof of Theorem 2.3. This establishes the first claim of the theorem.

Under Assumption 2.6, the demand for bundle (1,1) can be written as (2.29). It then follows that for δ_{-j} sufficiently large, one may define

$$\tilde{\Phi}_{(1,1)}(x_j^{(2)}, p_j, \delta_j) \equiv \phi_{(1,1)}(x^{(2)}, p, \delta) = \int 1\{x_j^{(2)\prime} b^{(2)} + ap_j + \Delta > -\delta_j\} f_\theta(b^{(2)}, a, \Delta) d\theta \ .$$

Let $l = (1,1)$, $w \equiv -(x_j^{(2)}, p_j, 1)/\|(x_j^{(2)}, p_j, 1)\|$ and $u \equiv \delta_j/\|(x_j^{(2)}, p_j, 1)\|$. Define Φ_l as in (2.23). Arguing as in (5.2) and taking a derivative with respect to u, we have $\partial\Phi_l(w,u)/\partial u = \mathcal{R}[f_\theta](w,u)$. However since w is a vector of normalized covariates including a constant, it takes values in $\mathbb{H}_-$, where $\mathbb{H}_-$ is the hemisphere in $\mathbb{R}^{d_\theta}$ defined by $\mathbb{H}_- \equiv \{w = (w_1, w_2, \cdots, w_{d_\theta}) \in \mathbb{S}^{d_\theta - 1} : w_{d_\theta} \leq 0\}$. Therefore, $\partial\Phi_l(w,u)/\partial u$ is well-defined for all $(w,u) \in \mathbb{H}_- \times \mathbb{R}$. Note that the Radon transform satisfies the symmetry $\mathcal{R}[f_\theta](w,u) = \mathcal{R}[f_\theta](-w,-u)$. Hence, for any $w \in \mathbb{S}^{d_\theta - 1} \setminus \mathbb{H}_-$, we have $\partial\Phi_l(w',u')/\partial u'|_{(w',u')=(w,u)} = \mathcal{R}[f_\theta](w,u) = \mathcal{R}[f_\theta](-w,-u) = \partial\Phi_l(w',u')/\partial u'|_{(w',u')=(-w,-u)}$. Since $(-w,-u) \in \mathbb{H}_- \times \mathbb{R}$, this means that $\partial\Phi_l(w,u)/\partial u$ is also well-defined on $\mathbb{S}^{d_\theta - 1} \times \mathbb{R}$. The conclusion of the theorem then follows from the injectivity of the Radon transform and Φ_l being identified. ∎

Proof of Theorem 2.5. Arguing as in the Proof of Theorem 2.4, Assumptions 2.1-2.3 and Theorem 2.2 ensure identification of ϕ. The aggregate demand on (0,1) is then given by

$$\begin{aligned}
\phi_{(0,1)}(x^{(2)}, p, \delta) = &\int 1\{x_2^{(2)\prime} b^{(2)} + ap_2 > -\delta_2\} \\
&\times 1\{(x_2^{(2)} - x_1^{(2)})' b^{(2)} + a(p_2 - p_1) > -(\delta_2 - \delta_1)\} 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + \Delta < -\delta_1\} f_\theta(b^{(2)}, a, \Delta) d\theta \ ,
\end{aligned}$$

By Assumption 2.8, $\Delta \geq 0, a.s.$. This implies that

$$\phi_{(0,1)}(x^{(2)}, p, \delta) = \int 1\{x_2^{(2)\prime} b^{(2)} + ap_2 > -\delta_2\} 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + \Delta < -\delta_1\} f_\theta(b^{(2)}, a, \Delta) d\theta$$

Without loss of generality, suppose Assumption 2.7 holds with $j = 1$ and $-j = 2$. Then, $\tilde{\Phi}_{(0,1)}$ in (2.33) is well-defined. Define Φ_l as in (2.23) with $l = (0,1)$, $w = (x_1^{(2)}, p_1, 1)/\|(x_1^{(2)}, p_1, 1)\|$ and $u = -\delta_1/\|(x_1^{(2)}, p_1, 1)$. The rest of the proof is then the same as the proof of Theorem 2.4 (ii). ∎

Proof of Theorem 2.6. First, let $\tilde{\mathbb{L}} = \{(2,0),(2,1)\}$. By Condition 2.1 and Lemma 1, Assumption 2.2 is satisfied. By Assumptions 2.1-2.3 and Theorem 2.2, ψ is identified. This implies that the aggregate demand $\{\phi_l, l = (2,0),(2,1)\}$ is identified by (2.14). Second, take $\tilde{\mathbb{L}} = \{(0,0),(0,1)\}$. Then by the same argument, the aggregate demand $\{\phi_l, l = (0,0),(0,1)\}$ is identified as well. Again by Condition 2.1, we may let δ_2 sufficiently large so that $\tilde{\Phi}_l, l \in \{(0,1),(2,0),(2,1)\}$ in (2.34)-(2.36) are well defined. For each $l \in \{(0,1),(2,0),(2,1)\}$, define Φ_l as in (2.23) with $w \equiv -(x_1^{(2)}, p_1, 1)/\|(x_1^{(2)}, p_1, 1)\|$ and $u \equiv \delta_1/\|(x_1^{(2)}, p_1, 1)\|$. Then, it follows that

$$\frac{\partial \Phi_l(w,u)}{\partial u} = \mathcal{R}[f_l](w,u), \quad l \in \{(0,1),(2,0),(2,1)\},$$

where $f_{(0,1)} = f_{(\beta^{(2)},\alpha,\Delta_{(1,1)})}, f_{(2,0)} = f_{(\beta^{(2)},\alpha,\Delta_{(2,0)})}$, and $f_{(2,1)} = f_{(\beta^{(2)},\alpha,\Delta_{(2,1)}-\Delta_{(1,1)})}$. By inverting the Radon transform, the random coefficient densities $f_l, l \in \{(0,1),(2,0),(2,1)\}$ are identified. It remains to construct the joint density from these marginal densities. In what follows, the arguments are made conditional on $(\beta^{(2)}, \alpha)$ unless otherwise noted. From the previous step, the marginal densities of $\Delta_{(2,1)} - \Delta_{(1,1)}$ and $\Delta_{(1,1)}$ are identified. Further, we note that $\Delta_{(2,1)} - \Delta_{(1,1)}$ is a convolution of $\Delta_{(2,1)}$ and $-\Delta_{(1,1)}$. By Assumption 2.9 (ii), Proposition 8 of Carrasco and Florens (2010) applies. Hence, the marginal density of $\Delta_{(2,1)}$ is identified. By Assumption 2.9 (i), $\Delta_{(1,1)} \perp \Delta_{(2,0)} \perp \Delta_{(2,1)}$ conditional on $(\beta^{(2)}, \alpha)$, and each of the marginal densities was identified in the previous step. Therefore, the joint density $f_{(\Delta_{(1,1)},\Delta_{(2,0)},\Delta_{(2,1)})|(\beta^{(2)},\alpha)}$ is identified as the product of the marginal densities. Finally, since the density of $(\beta^{(2)}, \alpha)$ can also be identified from any of $f_l, l \in \{(0,1),(2,0),(2,1)\}$, we may identify the joint density f_θ as $f_\theta = f_{(\Delta_{(1,1)},\Delta_{(2,0)},\Delta_{(2,1)})|(\beta^{(2)},\alpha)} f_{(\beta^{(2)},\alpha)}$. This establishes the claim of the theorem. ∎