Courant Research Centre
'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 224

Nonparametric Identification of Random Coefficients in Endogenous and Heterogeneous Aggregate Demand

Fabian Dunker, Stefan Hoderlein, Hiroaki Kaido

April 2017

Platz der Göttinger Sieben 5 · 37073 Goettingen · Germany
Phone: +49-(0)551-3921660 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Nonparametric Identification of Random Coefficients in Endogenous and Heterogeneous Aggregate Demand Models

Fabian Dunker* Stefan Hoderlein† Hiroaki Kaido‡

University of Goettingen Boston College Boston University

February 13, 2017

Abstract

This paper studies nonparametric identification in market level demand models for differentiated products with heterogeneous consumers. We consider a general class of models that allows for the individual specific coefficients to vary continuously across the population and give conditions under which the density of these coefficients, and hence also functionals such as welfare measures, is identified. Building on earlier work by Berry and Haile (2013), 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) the variation of the product characteristics across markets that is exogenous to the individual heterogeneity. We further show that two leading models, the BLP-model (Berry, Levinsohn, and Pakes, 1995) and the pure characteristics model (Berry and Pakes, 2007), require considerably different conditions on the support of the product characteristics.

*CRC Poverty, Equity and Growth, University of Goettingen, Humboldtallee 3, 37073 Goettingen, Germany

†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, Kei Hirano, Arthur Lewbel, Marc Rysman, and Elie Tamer for their helpful comments. We also thank seminar and conference participants at Harvard, 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 SFB 823 (projects C1, C4) and from the Ministry of Education and Cultural Affairs of Lower Saxony in the project Reducing Poverty Risk. 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, urban or education economics, 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), market-level demand models 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 empiricist 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

[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.

of a heterogeneous utility index are nonparametrically identified.

Our work complements this line of work in that we follow more closely the original BLP 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). For example, consider a change in the characteristics of a good. 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 the random coefficient density allows the researcher to calculate the distribution of the welfare effects. This allows one to answer various questions. For example, one may investigate whether the change gives rise to a Pareto improvement. This is possible because, with the distribution of the random coefficients being identified, one can track each individual's welfare before and after the change. If a change in one of the product characteristics is not Pareto improving, one can also calculate the proportion of individuals who would benefit from the change and therefore prefers the product with new characteristics.[2] Identification of the random coefficient distribution allows one to conduct various types of welfare analysis that are not possible by only identifying the demand function. Our focus therefore will be on the set of conditions under which one can uniquely identify the random coefficient distribution from the observed demand.

Naturally, Identification will depend crucially on the specific model at hand. As it turns out, there are important differences between the classical BLP and the pure characteristics model (see Berry and Pakes (2007), PCM henceforth) that stem from the presence of an alternative, individual and market specific error, typically assumed to be logistically distributed and hence called "logit error" in the following. A lucid discussion about the pros and cons of both approaches can be found in Berry and Pakes (2007). One advantage of the PCM we would like to emphasize at this point is that it is well-suited for the analysis of welfare changes when a new product with a particular characteristic is introduced to the market. Moreover, the pure characteristics model also predicts a reasonable substitution pattern when the number of products is large, while the BLP-type model may give counter-intuitive predictions. In

[2]Note that a simultaneous change in the product characteristic and price is allowed. Hence, one can also investigate how much price change can be made to compensate for a change (e.g. downgrading of a feature) in one of the product characteristics to let a certain fraction of individuals receive a non-negative utility change, i.e. $P(\Delta U_{ijt} \geq 0) \geq \tau$ for some prespecified $\tau \in [0, 1]$, where ΔU_{ijt} denotes the utility change.

addition to these important economic differences, the identification strategies including the required assumptions also differ significantly across the two models. In particular, in the BLP model, one needs to rely on an identification at infinity argument to isolate the unobservable for each product. In remarkable contrast, in the PCM one does not require such an argument (and therefore does not have to employ some restrictive assumptions). Instead, in the PCM, we demonstrate that one may combine demand on products across different markets to construct a function that depends on the random coefficients through a single index so that we can recover the distribution of unobserved heterogeneity without relying on identification at infinity. We call this construction *marginalization (or aggregation) of demand*. This is possible due to the unique structure of the PCM in which only the product characteristics (but not the tastes for products) determine the demand. To our knowledge, this identification strategy is novel.

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.

Another 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 Gowrisankaran and Rysman (2012), Armstrong (2013) and Moon, Shum, and Weidner (2013). Less closely related is the literature on hedonic models, see Heckman, Matzkin and Nesheim (2010), and references therein.

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 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). Finally, the class of models we consider is related but differs from the mixed logit model (without endogeneity) analyzed by Fox, Kim, Ryan, and Bajari (2012) who established the identification of the distribution of the random coefficients from micro-level data, while maintaining the logit assumption on the tastes for products. Our focus here is on market-level models with endogeneity with the main goal being the identification of the distribution of all random coefficients without any parametric assumption. As such, our identification strategy differs significantly from theirs.

Structure of the paper: The second section lays out preliminaries we require for our main result: We first introduce the class of models and detail the structure of our two main setups. Still in the same section, for completeness we quickly recapitulate the results of Berry and Haile (2013) concerning the identification of structural demands, adapted to our setup. The third section contains the key novel result in this paper, the nonparametric (point-)identification of the distribution of random coefficients in the class of discrete choice demand model with IV type endogeneity, which includes the BLP and PCM models. The fourth section contains various extensions: We discuss the identification in the bundles case, including how the structural demand identification results of Berry and Haile (2013) have to be adapted, but again focusing on the random coefficients density. As another set of extensions, we discuss the multiple units case. Finally, we discuss how full independence assumptions may be utilized to increase the strength of identification, in particular in the identification of structural demands. The fifth section discusses estimation. The objective here is twofold, first we sketch how a nonparametric sample counterparts estimator that utilizes the insights of the identification sections could be constructed, and we propose a simple parametric estimator for the bundles model which we believe to be relevant for applications. We end with an outlook.

2 Preliminaries

2.1 Model

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 individuals by i, 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 individual i in market t, the (indirect) utility from consuming good j depends on its (log) price P_{jt}, a vector of observable characteristics $X_{jt} \in \mathbb{R}^{d_X}$, and an unobservable scalar characteristic $\Xi_{jt} \in \mathbb{R}$. We model the utility from consuming good j using the linear random coefficient specification:

$$U^*_{ijt} \equiv X'_{jt}\beta_{it} + \alpha_{it}P_{jt} + \Xi_{jt} + \sigma_\epsilon \epsilon_{ijt},\ j = 1, \cdots, J\ , \tag{2.1}$$

where $(\alpha_{it}, \beta_{it})' \in \mathbb{R}^{d_X+1}$ is a vector of random coefficients representing the tastes for the product characteristics. For each j, ϵ_{ijt} represents the "taste for the product" itself. Following Berry and Pakes (2007), we consider a class of general market-level demand models that nests models with tastes for products ($\sigma_\epsilon = 1$) and without tastes for products ($\sigma_\epsilon = 0$). The models with tastes for products include the random coefficient logit model used in BLP, in which case $\sigma_\epsilon = 1$ and $\epsilon_{ijt}, j = 1, \cdots, J$ are i.i.d. Type-I extreme value random variables. When $\sigma_\epsilon = 0$, the model is called the *pure characteristic model* (PCM). The two models are known to have different theoretical properties. For example, the BLP model predicts that even with a large number of products, the mark-up remains positive implying there is always an incentive to develop a new product. As the number of new products grows, each individual's utility tends to infinity. On the other hand, in PCM, the model approaches competitive equilibrium and the incentive to develop a new product diminishes as the number of products increases.[3] As we will show below, the two models also differ in terms of empirical contents.

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).

We think of a large sample of individuals as *iid* copies of this population model. The random coefficients $\theta_{it} \equiv (\alpha_{it}, \beta_{it}, \epsilon_{i1t}, \cdots, \epsilon_{iJt})'$ 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. These coefficients are assumed to follow a distribution with a density function f_θ with respect to Lebesgue measure, i.e., be continuously distributed.[4] This density is assumed to be common across markets, and is therefore not indexed by t. As we will show, an important aspect of our identification argument is that, once the demand function is identified, one may recover Ξ_t from the market shares and other product characteristics (X_t, P_t). Then, by creating exogenous variations in the product characteristics and exploiting the linear random coefficients structure, one may trace out the distribution f_θ of the preference that is common across markets. We note that we can allow for the coefficients $(\alpha_{it}, \beta_{it})$ 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 two examples. The first example is the setting of the BLP and pure characteristics models, where individuals choose a single good out of multiple products, while the second is about the demand for bundles.

Example 1 (Multinomial choice)**.** *Each individual chooses the product that maximizes her*

[3]See Berry and Pakes (2007) for more details.

[4]This assumption is not crucial but made for the ease of exposition. Our main identification results (Theorems 3.1 and 3.2) hold for any Borel measure.

utility out of $J \in \mathbb{N}$ products. Hence, product j is chosen if

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

The demand for good j in market t is obtained by aggregating the individual demand with respect to the distribution of individual preferences.

$$\varphi_j(X_t, P_t, \Xi_t) = \int 1\{X'_{jt}b + aP_{jt} + \sigma_\epsilon e_j > -\Xi_{jt}\} 1\{(X_{jt} - X_{1t})'b + a(P_{jt} - P_{1t}) + \sigma_\epsilon(e_j - e_1) > -(\Xi_{jt} - \Xi_{1t})\} \cdots 1\{(X_{jt} - X_{Jt})'b + a(P_{jt} - P_{Jt}) + \sigma_\epsilon(e_j - e_J) > -(\Xi_{jt} - \Xi_{Jt})\} f_\theta(b, a, e) d\theta \ , \tag{2.3}$$

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

$$\varphi_0(X_t, P_t, \Xi_t) = \int 1\{X'_{1t}b + aP_{1t} + \sigma_\epsilon e_1 < -\Xi_{1t}\} \cdots 1\{X'_{Jt}b + aP_{Jt} + \sigma_\epsilon e_J < -\Xi_{Jt}\} f_\theta(b, a, e) d\theta \ , \tag{2.4}$$

where $(b, a, e_1, \cdots, e_J)$ are placeholders for the random coefficients $\theta_{it} = (\beta_{it}, \alpha_{it}, \epsilon_{i1t}, \cdots, \epsilon_{iJt})$. The researcher then observes the market shares of products $S_{lt} = \varphi_l(X_t, P_t, \Xi_t), l \in \mathbb{L}$, where $\mathbb{L} = \{0, 1, \cdots, J\}$.

The second class of examples considers discrete choice, but allows for the choice of bundles.

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_{it} if the two goods are consumed simultaneously. Here, Δ_{it} is also allowed to vary across individuals. The utility $U^*_{i,(Y_1,Y_2),t}$ from each bundle is therefore specified as follows:

$$\begin{aligned}
&U^*_{i,(0,0),t} = 0, \\
&U^*_{i,(1,0),t} = X'_{1t}\beta_{it} + \alpha_{it}P_{1t} + \Xi_{1t} + \sigma_\epsilon \epsilon_{i1t}, \quad U^*_{i,(0,1),t} = X'_{2t}\beta_{it} + \alpha_{it}P_{2t} + \Xi_{2t} + \sigma_\epsilon \epsilon_{i2t}, \\
&U^*_{i,(1,1),t} = X'_{1t}\beta_{it} + X'_{2t}\beta_{it} + \alpha_{it}P_{1t} + \alpha_{it}P_{2t} + \Xi_{1t} + \Xi_{2t} + \sigma_\epsilon \epsilon_{i1t} + \sigma_\epsilon \epsilon_{i2t} + \Delta_{it} \ ,
\end{aligned} \tag{2.5}$$

Each individual chooses a bundle that maximizes her utility. Hence, bundle (y_1, y_2) is chosen when $U^*_{i,(y_1,y_2),t} > U^*_{i,(y_1',y_2'),t}$ for all $(y_1', y_2') \neq (y_1, y_2)$. For example, bundle $(1,0)$ is chosen if

$$X_{1t}'\beta_{it}+\alpha_{it}P_{1t}+\Xi_{1t}+\sigma_\epsilon\epsilon_{i1t} > 0, \text{ and } X_{1t}'\beta_{it}+\alpha_{it}P_{1t}+\Xi_{1t}+\sigma_\epsilon\epsilon_{i1t} > X_{2t}'\beta_{it}+\alpha_{it}P_{2t}+\Xi_{2t}+\sigma_\epsilon\epsilon_{i2t}, \text{ and}$$
$$X_{1t}'\beta_{it}+\alpha_{it}P_{1t}+\Xi_{1t}+\sigma_\epsilon\epsilon_{i1t} > X_{1t}'\beta_{it}+\alpha_{it}P_{1t}+\Xi_{1t}+\sigma_\epsilon\epsilon_{i1t}+X_{2t}'\beta_{it}+\alpha_{it}P_{2t}+\Xi_{2t}+\sigma_\epsilon\epsilon_{i2t}+\Delta_{it}\ . \tag{2.6}$$

Suppose the random coefficients $\theta_{it} = (\beta_{it}', \alpha_{it}, \Delta_{it}, \epsilon_{i1t}, \epsilon_{i2t})$ 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:

$$\varphi_{(1,0)}(X_t, P_t, \Xi_t) = \int 1\{X_{1t}'b+aP_{1t}+\sigma_\epsilon e_1 > -\Xi_{1t}\}1\{(X_{1t}-X_{2t})'b+a(P_{1t}-P_{2t})+\sigma_\epsilon(e_1-e_2) > \Xi_{2t}-\Xi_{1t}\}$$
$$\times 1\{X_{2t}'b + aP_{2t} + \sigma_\epsilon e_2 + \Delta < -\Xi_{2t}\} f_\theta(b, a, \Delta, e) d\theta\ . \tag{2.7}$$

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

In Examples 2, 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. Examples 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 Example 2 (and Example 3 in Section 4), we henceforth tacitly assume to have access to such data in principle.

2.2 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 $X_{jt} = (X_{jt}^{(1)}, X_{jt}^{(2)}) \in \mathbb{R} \times \mathbb{R}^{d_X - 1}$, and make the following assumption.

Assumption 2.1. *The coefficient $\beta_{ij}^{(1)}$ on $X_{jt}^{(1)}$ is non-random for all j and is normalized to 1.*

Assumption 2.1 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 $\beta_{ij}^{(1)} = 1$ for all

j. Under Assumption 2.1, the utility for product j can be written as $U^*_{jt} = X^{(2)'}_{jt}\beta_{ij}{}^{(2)} + \alpha_{ij}P_{jt} + \sigma_\epsilon\epsilon_{ijt} + D_{jt}$, where $D_{jt} \equiv X^{(1)}_{jt} + \Xi_{jt}$ 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^*_{ijt} 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 what follows, we redefine the aggregate demand as a function of $(X^{(2)}_t, P_t, D_t)$ instead of (X_t, P_t, Ξ_t) by $\phi(X^{(2)}_t, P_t, D_t) \equiv \varphi(X_t, P_t, \Xi_t)$, where $X_t = (X^{(1)}_t, X^{(2)}_t)$ and $D_t = \Xi_t + X^{(1)}_t$ and make the following assumption

Assumption 2.2. *For some subset $\tilde{\mathbb{L}}$ of $\mathbb{L}$ whose cardinality is J, there exists a unique function $\psi : \mathbb{R}^{J\times(d_X-1)} \times \mathbb{R}^J \times \mathbb{R}^J \to \mathbb{R}^J$ such that $D_{jt} = \psi_j(X^{(2)}_t, P_t, \tilde{S}_t)$ 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

$$\Xi_{jt} = \psi_j(X^{(2)}_t, P_t, \tilde{S}_t) - X^{(1)}_{jt}. \tag{2.8}$$

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. 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.*

In Example 2, one may employ an alternative inversion strategy to obtain ψ in (2.8) using only subsystems of demand such as $\tilde{\mathbb{L}} = \{(1,0),(1,1)\}$ or $\tilde{\mathbb{L}} = \{(0,0),(0,1)\}$. We defer details on this case to Section 4.

The inverted system in (2.8), 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 $E[|B(X^{(2)}_t, P_t, \tilde{S}_t)|] < \infty$, it holds that*

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

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. In Appendix C, 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 $X_{jt}^{(1)} + \Xi_{jt}$. Where $X_{jt}^{(1)}$ has a close analogy to a dependent variable in nonparametric IV.

Given Assumption 2.3 and (2.8), the unknown function ψ can be identified through the following conditional moment restrictions:

$$E[\psi_j(X_t^{(2)}, P_t, S_t) - X_{jt}^{(1)} | Z_t, X_t] = 0, \quad j = 1, \cdots, J. \tag{2.9}$$

We here state this result as a theorem.

Theorem 2.1. *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 (Multinomial choice, continued)**.** *Recall that ψ is a unique function such that*

$$S_{jt} = \phi_j(X_t^{(2)}, P_t, D_t), \; j = 1, \cdots, J \quad \Leftrightarrow \quad \Xi_{jt} = \psi_j(X_t^{(2)}, P_t, \tilde{S}_t) - X_{jt}^{(1)}, \quad j = 1, \cdots, J, \tag{2.10}$$

where $\tilde{S}_t = (S_{1t}, \cdots, S_{Jt})$. Hence, the structural demand $(\phi_1, \cdots, \phi_J)$ is identified by Theorem 2.1 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 $\tilde{\mathbb{L}} = \{(1,0), (1,1)\}$. ψ is then a unique function such that*

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

where $\tilde{S}_t = (S_{(1,0),t}, S_{(1,1),t})$. Theorem 2.1 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, one may also identify the rest of the subcomponents by taking $\tilde{\mathbb{L}} = \{(0,0), (0,1)\}$ and applying Theorem 2.1 again.

3 Identification of the Random Coefficient Density

This section contains the main innovation in this paper: We establish that the density of random coefficients in the market-level demand models is nonparametrically identified. 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, we construct a function $\Phi(w,u)$ such that

$$\frac{\partial \Phi(w,u)}{\partial u} = \mathcal{R}[f](w,u) \ , \tag{3.1}$$

where f is the density of interest, w is a vector in $\mathbb{R}^q$ (with q the dimension of the random coefficients), normalized to have unit length, and $u \in \mathbb{R}$ is a scalar. In what follows, we let $\mathbb{S}^q \equiv \{v \in \mathbb{R}^q : \|v\| = 1\}$ denote the unit sphere in $\mathbb{R}^q$. $\mathcal{R}$ is the Radon transform defined pointwise by

$$\mathcal{R}[f](w,u) = \int_{P_{w,u}} f(v) d\mu_{w,u}(v). \tag{3.2}$$

where $P_{w,u}$ denotes the hyperplane $\{v \in \mathbb{R}^q : v'w = u\}$, 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 Beran, Feuerverger, and Hall (1996), Hoderlein, Klemelä, and Mammen (2010), and Gautier and Hoderlein (2013).

Throughout, we maintain the following assumption.

Assumption 3.1. *(i) For all $j \in \{1, \cdots, J\}$, $(X_{jt}^{(2)}, P_{jt}, D_{jt})$ are absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^{d_X - 1} \times \mathbb{R} \times \mathbb{R}$; (ii) the random coefficients θ are independent of (X_t, P_t, D_t).*

Assumption 3.1 (i) requires that $(X_{jt}^{(2)}, P_{jt}, D_{jt})$ are continuously distributed for all j. By Assumption 3.1 (ii), 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.

Before proceeding further, we overview our identification strategy in relation to the key differences between the BLP and pure characteristics models. Heuristically, for a given $(w,u) \in \mathbb{S}^q \times \mathbb{R}$, the Radon transform aggregates individuals whose coefficients are on the hyperplane $P_{w,u}$. For each (w,u), we relate this aggregate value to a feature of the demand with a specific product characteristics. By varying (w,u) and inverting the map $\mathcal{R}$ in (3.1), we may then recover the distribution of the random coefficients. A key step in this identification argument is the construction of a function Φ satisfying (3.1). The two demand models suggest different strategies to construct Φ. In the BLP model, we construct Φ for each product j and recover the joint distribution of the coefficients $(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{ijt})$. We take this approach because the presence of the tastes for products requires us to isolate the demand for each product from the rest. On the other hand, the pure characteristics model does not require such an approach. Furthermore, both models allow the researcher to combine demand across different markets to construct Φ.

3.1 BLP model

Throughout this section, we let $\sigma_\epsilon = 1$.[5] 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:

$$\phi_j(x^{(2)}, p, \delta) = \int 1\{x_j^{(2)\prime} b^{(2)} + ap_j + e_j > -\delta_j\} 1\{(x_j^{(2)} - x_1^{(2)})' b^{(2)} + a(p_j - p_1) + (e_j - e_1) > -(\delta_j - \delta_1)\}$$
$$\cdots 1\{(x_j^{(2)} - x_J^{(2)})' b^{(2)} + a(p_j - p_J) + (e_j - e_J) > -(\delta_j - \delta_J)\} f_\theta(b^{(2)}, a, e) d\theta \ . \quad (3.3)$$

Suppose the vertical characteristics $\{D_{kt}, k \neq j\}$ (for products other than j) have a large enough support so that $(X_{jt}^{(2)} - X_{kt}^{(2)})' \beta_{it}^{(2)} + \alpha_{it}(P_{jt} - P_{kt}) + (\epsilon_{ijt} - \epsilon_{ikt}) - D_{jt} > D_{kt}$ for all $k \neq j$ for some values of $D_{kt}, k \neq j$. The demand for good j for such values of $D_{kt}, k \neq j$ is then

$$\begin{aligned} \tilde{\Phi}_j(x_j^{(2)}, p_j, \delta_j) &= \lim_{\delta_1, \ldots, \delta_{j-1}, \delta_{j+1}, \ldots, \delta_J \to -\infty} \phi_j(x^{(2)}, p, \delta) \\ &= \int 1\{x_j^{(2)\prime} b^{(2)} + ap_j + e_j < -\delta_j\} f_{\vartheta_j}(b^{(2)}, a, e_j) d\vartheta_j, \end{aligned} \quad (3.4)$$

where f_{ϑ_j} is the joint density of the subvector $\vartheta_{ijt} \equiv (\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{ijt})$ of the random coefficients. Let $w \equiv (x_j^{(2)}, p_j, 1)/\|(x_j^{(2)}, p_j, 1)\|$ and $u \equiv \delta_j/\|(x_j^{(2)}, p_j, 1)\|$. Define

$$\begin{aligned} \Phi(w, u) &\equiv \tilde{\Phi}_j\left(\frac{x_j^{(2)}}{\|(x_j^{(2)}, p_j, 1)\|}, \frac{p_j}{\|(x_j^{(2)}, p_j, 1)\|}, \frac{\delta_j}{\|(x_j^{(2)}, p_j, 1)\|}\right) \\ &= \tilde{\Phi}_j(x_j^{(2)}, p_j, \delta_j), \quad (x_j^{(2)}, p_j, \delta_j) \in \operatorname{supp}(X_{jt}^{(2)}, P_{jt}, D_{jt}), \end{aligned} \quad (3.5)$$

[5]Here, the scale of the taste for product ϵ_{ijt} is normalized relative to the scale of $X_{jt}^{(1)}$ as we set the coefficient on $X_{jt}^{(1)}$ to 1 in Assumption 2.1.

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

$$\Phi(w,u) = -\int 1\{w'\vartheta_j < -u\} f_{\vartheta_j}(b^{(2)}, a, e_j) d\vartheta_j$$

$$= -\int_{-\infty}^{-u} \int_{P_{w,r}} f_{\vartheta_j}(b^{(2)}, a, e_j) d\mu_{w,r}(b^{(2)}, a, e_j) dr = -\int_{-\infty}^{-u} \mathcal{R}[f_{\vartheta_j}](w,r) dr \ , \quad (3.6)$$

Hence, by taking a derivative with respect to u, we may relate Φ to the random coefficient density through the Radon transform:

$$\frac{\partial \Phi(w,u)}{\partial u} = \mathcal{R}[f_{\vartheta_j}](w,u). \quad (3.7)$$

Note that since the structural demand ϕ is identified by Theorem 2.1, Φ is nonparametrically identified as well. Hence, Eq. (3.7) gives an operator that maps the random coefficient density to an object identified by the moment condition studied in the previous section. To construct Φ described above and to invert the Radon transform, we formally make the following assumptions. Below, for each $1 \leq, j, k \leq J$, we let $V_{jk} = (X_{jt}^{(2)} - X_{kt}^{(2)})'\beta_{it}^{(2)} + \alpha_{it}(P_{jt} - P_{kt}) + (\epsilon_{ijt} - \epsilon_{ikt}) - D_{jt}$ and make the following assumptions on the support of the product characteristics.[6]

Assumption 3.2. *Let $\mathcal{J}$ be a nonempty subset of $\{1, \cdots, J\}$. For each $j \in \mathcal{J}$,* $\operatorname{supp}(V_{jk}, k \neq j) \subset \operatorname{supp}(D_{kt}, k \neq j)$.

Assumption 3.2 requires that one may vary the vertical characteristics of the alternative products $\{D_{kt}, k \neq j\}$ on a large enough support so that the demand for product j is determined through its choice between product j and the outside good. This identification argument therefore uses a "thin" (lower-dimensional) subset of the support of the covariates, which is due to the presence of the tastes for products. This is in remarkable contrast with the identification of the random coefficients density in the PCM which does not rely on thin sets.

Assumption 3.3. *One of the following conditions hold*

(i) $\bigcup_{j \in \mathcal{J}} \operatorname{supp}(X_{jt}^{(2)}, P_{jt}, D_{jt})$ *has full support in* $\mathbb{R}^{d_X - 1} \times \mathbb{R} \times \mathbb{R}$.

(ii) $\bigcup_{j \in \mathcal{J}} \operatorname{supp}(X_{jt}^{(2)}, P_{jt})$ *contains an open ball* $B_{\mathcal{J}} \subset \mathbb{R}^{d_X - 1} \times \mathbb{R}$. *For every* $(x, p) \in B_{\mathcal{J}}$ *and every* $(b^{(2)}, a, e_j) \in \operatorname{supp}(\vartheta_j)$ *it holds that*

$$\left(x, p, -x'b^{(2)} - ap - e_j\right) \in \bigcup_{j \in \mathcal{J}} \operatorname{supp}\left(X_{jt}^{(2)}, P_{jt}, D_{jt}\right). \quad (3.8)$$

[6] V_{jk} is a random variable that varies across individuals and markets and hence should be denoted as V_{ijkt} in principle. For conciseness, we drop subscripts i and t below.

Furthermore, all the absolute moments of each component of θ_{it} *are finite, and for any fixed* $z \in \mathbb{R}_+$, $0 = \lim_{l\to\infty} \frac{z^l}{l!} E[(|\theta_{it}^{(1)}| + \cdots + |\theta_{it}^{(d_\theta)}|)^l]$.

Assumption 3.3 (i) is our benchmark assumption. Under this assumption, no restrictions on θ_{it} are necessary for identification. In fact, the identification strategy would be valid for arbitrary Borel measures and may also be applied to settings where θ_{it} does not have a density.[7] However, this large support assumption is stringent and may be violated by various product characteristics and prices used in practice. Hence, it should be viewed as a benchmark to understand what the model requires to identify the distribution f_θ of θ_{it} if one does not impose any restriction on it.

Assumption 3.3 (ii) is an alternative condition, which relaxes the support requirement significantly. Instead of a large support, it is enough for the product characteristics to have a properly combined support that contains a (possibly small) open ball $B_{\mathcal{J}}$ in it. This includes as a special case where a single product's characteristics $(X_{jt}^{(2)}, P_{jt})$ contains an open ball, which can be met in various applications. Even if such a product does not exist, identification of the random coefficient density is possible as long as the required support condition is met by combining the supports of multiple products belonging to $\mathcal{J}$. This means that our identification strategy may use variations of $(X_{jt}^{(2)}, P_{jt})$ across products. To illustrate, consider three products $J = 3$. If (D_{2t}, D_{3t}) have a large support in the sense of Assumption 3.2 ($\mathcal{J} = \{1\}$ in this case), identification of the random coefficient density is possible as long as the characteristics of good 1 contains an open ball. If all $\{D_{jt}\}_{j=1}^3$ jointly have a large support (this implies $\mathcal{J} = \{1, 2, 3\}$), our requirement on $(X_{jt}^{(2)}, P_{jt})$ becomes even milder as we only need to construct an open ball by combining the characteristics of all three products.

The condition in (3.8) allows for bounded support of $(X_{jt}^{(2)}, P_{jt})$. Further, if ϑ_j has a bounded support, Assumption 3.3 (ii) will allow for a bounded support of D_j. The price to pay for this relaxation of the support requirement is a regularity assumption on the moments of θ_{it}. This rules out heavy tailed distributions that are not determined by their moments. A sufficient, yet stronger than necessary, condition for this assumption is a compact support of f_θ. Under Assumption 3.3 (ii), the characteristic function $w \mapsto \varphi_{\vartheta_j}(tw)$ of ϑ_{ijt} (a key element of the Radon inversion) is analytic and thereby uniquely determined by its restriction to a non-empty full dimensional subset of its domain.[8] Hence, f_{ϑ_j} can be identified if one may vary $(X_{jt}^{(2)}, P_{jt})$ on a non-empty full dimensional subset.

[7]More precisely, the Radon transform $\mathcal{R}[f_{\vartheta_j}](w, u)$ gives f_{ϑ_j}'s integral along each hyperplane $P_{w,u} = \{v \in \mathbb{R}^{d_\theta} : v'w = u\}$ defined by the *angle* $w = (x_j^{(2)}, p_j, 1)/\|(x_j^{(2)}, p_j, 1)\|$ and *offset* $u = \delta_j/\|(x_j^{(2)}, p_j, 1)\|$. For recovering f_{ϑ_j} from its Radon transform, one needs exogenous variations in both. Our proof uses the fact that varying w over the hemisphere $\mathbb{H}_+ \equiv \{w = (w_1, w_2, \cdots, w_{d_{\vartheta_j}}) \in \mathbb{S}^{d_{\vartheta_j}-1} : w_{d_{\vartheta_j}} \geq 0\}$ and u over $\mathbb{R}$ suffices to recover f_{ϑ_j}.

[8]This type of moment condition on θ_{it} is common in the recent literature. See, for example, Hoderlein, Holzmann, and Meister (2014) and Masten (2014).

Under the conditions given in the theorem below, the Radon inversion identifies f_{ϑ_j}. If one is interested in the joint density of the coefficients on the product characteristics $(\beta_{it}^{(2)}, \alpha_{it})$, one may stop here as marginalizing f_{ϑ_j} gives the desired density. The joint distribution of the coefficients including the tastes for products can be identified under an additional independence assumption. We state this result in the following theorem.

Theorem 3.1. *Suppose Assumptions 2.1-3.3 hold. Suppose the conditional distribution of ϵ_{ijt} given $(\beta_{it}^{(2)}, \alpha_{it})$ is identical for all $j \in \mathcal{J}$. Then, (i) for each $j \in \mathcal{J}$, the density f_{ϑ_j} is identified, where $\vartheta_{ijt} = (\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{ijt})$; (ii) If, in addition, $\{\epsilon_{ijt}, j \in \mathcal{J}\}$ are independently distributed (across j) conditional on $(\beta_{it}^{(2)}, \alpha_{it})$, the joint density $f_{\theta_{\mathcal{J}}}$ of $\theta_{\mathcal{J}} = (\beta_{it}^{(2)}, \alpha_{it}, \{\epsilon_{ijt}\}_{j\in\mathcal{J}})$ is identified.*

An immediate corollary is the following.

Corollary 3.1. *Suppose Assumptions 2.1-3.3 hold. Suppose that $\{\epsilon_{ijt}\}_{j=1}^{J}$ is i.i.d. (across j) conditional on $(\beta_{it}^{(2)}, \alpha_{it})$. Then, the joint density f_θ of all random coefficients $\theta_{it} = (\beta_{it}^{(2)}, \alpha_{it}, \{\epsilon_{ijt}\}_{j=1}^{J})$ is identified.*

Several remarks are in order.

Remark 3.1. Theorems 2.1 and 3.1 shed light on the roles played by the key features of the BLP-type demand model: the invertibility of the demand system, instrumental variables, and the linear random coefficients specification. In Theorem 2.1, the invertibility and instrumental variables play key roles in identifying the demand. Once the demand is identified, one may "observe" the vector $(X_t^{(2)}, P_t, D_t)$ of product characteristics. This is possible because the invertibility of demand allows one to recover the unobserved product characteristics Ξ_t from the market shares S_t (together with other covariates). One may then vary $(X_t^{(2)}, P_t, D_t)$ across markets in a manner that is exogenous to the individual heterogeneity θ_{it}. Theorem 3.1 and Corollary 3.1 show that this exogenous variation combined with the linear random coefficients specification allows to trace out the distribution of θ_{it}.

Remark 3.2. The identical distribution assumption on the tastes for products in Theorem 3.1 is compatible with commonly used utility specifications and can also be relaxed at the cost of a stronger support condition on the product characteristics. In applications, it is often assumed that the utility of product j is

$$U_{ijt}^* = \beta_{it}^0 + \tilde{X}_{jt}'\beta_{it} + \alpha_{it}P_{jt} + \Xi_{jt} + \tilde{\epsilon}_{ijt}, \tag{3.9}$$

where $\tilde{X}_{jt}$ is a vector of non-constant product characteristics, β_{it}^0 is an individual specific intercept, which measures the utility difference between inside goods and the outside good, and $\tilde{\epsilon}_{ijt}$

is a mean zero error that follows the Type-I extreme value distribution. The requirement that $\epsilon_{ijt} = \beta_{it}^0 + \tilde{\epsilon}_{ijt}$ are i.i.d. across j (conditional on $(\beta_{it}, \alpha_{it})$) can be met if $\tilde{\epsilon}_{ijt}$ are i.i.d. across j.

If for each j, $(X_{jt}^{(2)}, P_{jt}, D_{jt})$ fulfills the support condition in Assumption 3.3 (i) or Assumption 3.3 (ii), one can drop the identical distribution assumption. This is because one can identify f_{ϑ_j} for all j by inverting the Radon transform in (3.7) repeatedly. This in turn implies that the distribution of ϵ_{ijt} conditional on $(\beta_{it}^{(2)}, \alpha_{it})$ is identified for each j. If the tastes for products $\{\epsilon_{ijt}\}_{j=1}^J$ are mutually independent (conditional on $(\beta_{it}^{(2)}, \alpha_{it})$), as is commonly assumed in BLP, the joint distribution f_θ of all coefficients is identified.

Finally, we comment on what an additional parametric assumption may add to our result. If one assumes that the tastes for products are i.i.d. and follows a parametric distribution, Eq (3.3) reduces to $\phi_j(x^{(2)}, p, \delta) = \int L(x^{(2)\prime(2)} + ap + \delta) f_{(\beta,\alpha)}(b, a) dbda$, for some function L, e.g. L is the logit function when $\{\epsilon_{ijt}\}$ follows a Type-I extreme value distribution. This type of integral equation is considered in Fox, Kim, Ryan, and Bajari (2012) in the context of individual-level demand model without endogeneity. Given that ϕ is identified, we believe that it is possible to extend their framework to the market-level demand model with endogeneity and identify $f_{(\beta,\alpha)}$ semiparametrically. This approach may allow us to relax some of the support conditions. To keep a tight focus on nonparametric identification, we leave this extension elsewhere.

Remark 3.3. Our identification result reveals the nature of the BLP-type demand model. A positive aspect of our result is that the preference is nonparametrically identified if one observes full dimensional variations in the consumers' choice sets (represented by $(X_{jt}^{(2)}, P_{jt}, D_{jt})$) across markets. The identifying power is quite strong, if the product characteristics jointly span a full support, i.e. $\bigcup_{j\in\mathcal{J}}(X_{jt}^{(2)}, P_{jt}, D_{jt}) = \mathbb{R}^{d_X-1} \times \mathbb{R} \times \mathbb{R}$. On the other hand, if the product characteristics have limited variations, the identifying power of the model on the distribution of preferences may be limited. In particular, identification is not achieved only with discrete covariates. Hence, for such settings, one needs to augment the model structure with a parametric specification. Another interesting direction would be to conduct partial identification analysis on functionals of f_θ, while imposing weak support restrictions. We leave this possibility for future research.

3.2 Pure Characteristics Demand Models

Throughout this section, we consider the following utility specification where each product's utility is fully determined by the tastes for the product characteristics:

$$U_{ijt}^* \equiv X_{jt}'\beta_{it} + \alpha_{it}P_{jt} + \Xi_{jt},\ j = 1, \cdots, J\ . \tag{3.10}$$

In other words, we set $\sigma_\epsilon = 0$ in (2.1). For this model, we employ a different, and arguably less restrictive, strategy from the one adopted in the previous section to construct Φ in (3.1). Below, we maintain Assumptions 2.1-2.3, which ensure the identification of demand by Theorem 2.1.

The demand for good j with the product characteristics (X_t, P_t, Ξ_t) is as given in (2.3) but with $\sigma_\epsilon = 0$. 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:

$$\begin{aligned}\phi_j(x^{(2)}, p, \delta) = \int & 1\{x_j^{(2)\prime} b^{(2)} + ap_j > -\delta_j\} 1\{(x_j^{(2)} - x_1^{(2)})' b^{(2)} + a(p_j - p_1) > -(\delta_j - \delta_1)\} \\ & \cdots 1\{(x_j^{(2)} - x_J^{(2)})' b^{(2)} + a(p_j - p_J) > -(\delta_j - \delta_J)\} f_\theta(b^{(2)}, a) d\theta \ . \quad (3.11)\end{aligned}$$

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:

$$(\acute{x}_j^{(2)} - \acute{x}_i^{(2)}, \acute{p}_j - \acute{p}_i, \acute{\delta}_j - \acute{\delta}_i) = -(x_j^{(2)} - x_i^{(2)}, p_j - p_i, \delta_j - \delta_i), \ \forall i \in \mathcal{J} \ , \quad (3.12)$$

$$(\acute{x}_i^{(2)}, \acute{p}_i, \acute{\delta}_i) = (x_i^{(2)}, p_i, \delta_i), \ \forall i \notin \mathcal{J} \ . \quad (3.13)$$

This map converts each product characteristic vector $(x^{(2)}, p, \delta)$ to another value $(\acute{x}^{(2)}, \acute{p}, \acute{\delta})$. Consider the composition $\phi_j \circ \mathcal{M}_{\mathcal{J}}(x^{(2)}, p, \delta)$. If $(\acute{x}^{(2)}, \acute{p}, \acute{\delta})$ is in the support, this corresponds to the demand of product j in some market (say t') with $(X_{t'}^{(2)}, P_{t'}, D_{t'}) = (\acute{x}^{(2)}, \acute{p}, \acute{\delta})$. We then define

$$\tilde{\Phi}_j(x_j^{(2)}, p_j, \delta_j) \equiv - \sum_{\mathcal{J} \subseteq \{1, \cdots J\} \setminus \{j\}} \phi_j \circ \mathcal{M}_{\mathcal{J}}(x^{(2)}, p, \delta) \ . \quad (3.14)$$

Eq (3.14) aggregates 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

$$\begin{aligned}\phi_1(x^{(2)}, p, \delta) = \int & 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\} \\ & \times 1\{(x_1^{(2)} - x_2^{(2)})' b^{(2)} + a(p_1 - p_2) < -(\delta_1 - \delta_2)\} f_\theta(b^{(2)}, a) d\theta \ . \quad (3.15)\end{aligned}$$

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

$$\begin{aligned}\tilde{\Phi}_1(x_1^{(2)}, p_1, \delta_1) &= -\phi_1 \circ \mathcal{M}_{\emptyset}(x^{(2)}, p, \delta) - \phi_1 \circ \mathcal{M}_{\{2\}}(x^{(2)}, p, \delta) \\ &= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\}\Big(1\{(x_1^{(2)} - x_2^{(2)})' b^{(2)} + a(p_1 - p_2) < -(\delta_1 - \delta_2)\} \\ &\qquad + 1\{(x_1^{(2)} - x_2^{(2)})' b^{(2)} + a(p_1 - p_2) > -(\delta_1 - \delta_2)\}\Big) f_\theta(b^{(2)}, a) d\theta \\ &= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 < -\delta_1\} f_\theta(b^{(2)}, a) d\theta \end{aligned} \tag{3.16}$$

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 a function $\tilde{\Phi}_1$ that depends only on product 1's characteristic $(x_1^{(2)}, p_1, \delta_1)$ through a single index in (3.16). This then allows us to trace out the random coefficients density by varying product 1's characteristic as done in the BLP model. Since the operation above yields a function that depends only on the characteristic of a single product, we call it *marginalization of demand*.[9]

Eq. (3.14) generalizes this argument to settings with $J \geq 2$. For the marginalization of demand to work, the product characteristic $(\acute{x}^{(2)}, \acute{p}, \acute{\delta}) = \mathcal{M}_{\mathcal{J}}(x^{(2)}, p, \delta)$ needs to be an observable value, meaning it must be in the support. Formally, a value of the product characteristic $(x^{(2)}, p, \delta) \in \text{supp}(X_t^{(2)}, P_t, D_t)$ is said to permit *marginalization of demand with respect to product* j if

$$\mathcal{M}_{\mathcal{J}}(x^{(2)}, p, \delta) \in \text{supp}(X_t^{(2)}, P_t, D_t), \ \forall \mathcal{J} \subseteq \{1, \cdots, J\} \setminus \{j\}. \tag{3.17}$$

As done in the BLP setting, we will only require that a rich enough set to recover f_θ can be constructed by combining the supports of multiple products' characteristics. Toward this end, for each $j \in \{1, \cdots, J\}$, let π_j be the projection map such that $(x_j^{(2)}, p_j, \delta_j) = \pi_j(x^{(2)}, p, \delta)$, and define the following sets:

$$\begin{aligned}\mathcal{H}_j \equiv \{&(x^{(2)}, p, \delta) \in \text{supp}(X_t^{(2)}, P_t, D_t) : \mathcal{M}_{\mathcal{J}}(x^{(2)}, p, \delta) \in \text{supp}(X_t^{(2)}, P_t, D_t), \\ &\text{for all } \mathcal{J} \subseteq \{1, \cdots, J\} \setminus \{j\}\},\end{aligned} \tag{3.18}$$

$$\mathcal{S}_j \equiv \{(x_j^{(2)}, p_j, \delta_j) \in \text{supp}(X_{jt}^{(2)}, P_{jt}, D_{jt}) : (x_j^{(2)}, p_j, \delta_j) = \pi_j(x^{(2)}, p, \delta) \ \text{ for some } (x^{(2)}, p, \delta) \in \mathcal{H}_j\}. \tag{3.19}$$

In words, $\mathcal{H}_j$ is the set of the entire product characteristic vectors for which marginalization with respect to product j is permitted. $\mathcal{S}_j$ is the coordinate projection of $\mathcal{H}_j$ onto the space of product j's characteristics. We then make the following assumption.

[9] Note however that this marginalized demand still depends on the joint distribution of the entire random coefficient vector.

Assumption 3.4. *One of the following conditions hold:*

(i) $\bigcup_{j=1}^{J} \mathcal{S}_j = \mathbb{R}^{d_X - 1} \times \mathbb{R} \times \mathbb{R}$*;*

(ii) $\bigcup_{j=1}^{J} \mathcal{S}_j = \mathbb{E} \times \mathbb{D}$*, where* $\mathbb{E}$ *contains an open ball* $B \subset \mathbb{R}^{d_X-1} \times \mathbb{R}$*, and* $\mathbb{D} \subseteq \mathbb{R}$*. For every* $(x, p) \in B$ *and every* $(b^{(2)}, a) \in \operatorname{supp}(\theta_{it})$*, it holds that* $(x, p, -x'b^{(2)} - ap) \in \bigcup_{j=1}^{J} \mathcal{S}_j$*.*

Furthermore, all the absolute moments of each component of θ_{it} *are finite, and for any fixed* $z \in \mathbb{R}_+$*,* $0 = \lim_{l \to \infty} \frac{z^l}{l!}(E[|\theta_{it}^{(1)}|^l] + \cdots + E[|\theta_{it}^{(d_\theta)}|^l])$*.*

The idea behind Assumption 3.4 is as follows. For the moment, suppose we don't impose any moment condition on the random coefficient density. Also, fix a benchmark product j. For any $(x_j^{(2)}, p_j, \delta_j) \in \mathcal{S}_j$, one may find a vector $(x^{(2)}, p, \delta)$ of all product characteristics for which marginalization of demand is allowed. Then, one would wish to vary $(x_j^{(2)}, p_j, \delta_j)$ to trace out the random coefficient density. This is possible, of course, if $\mathcal{S}_j = \mathbb{R}^{d_X-1} \times \mathbb{R} \times \mathbb{R}$, meaning that marginalization is possible everywhere with respect to product j. However, this assumption may be too strong in empirical applications. One may not be able to find any single product, for which this condition is satisfied. Assumption 3.4 (i) relaxes this requirement substantially using the structure of the model. Observe that the identification argument is symmetric across products because only the characteristics matter. Hence, the argument is valid as long as, for each $(\mathbf{x}^{(2)}, \mathbf{p}, \mathbf{d}) \in \mathbb{R}^{d_X-1} \times \mathbb{R} \times \mathbb{R}$, one can find *some* product for which marginalization is permitted. This is the reason why it is enough to "patch" $\mathcal{S}_j$s together to $\mathbb{R}^{d_X-1} \times \mathbb{R} \times \mathbb{R}$ in Assumption 3.4 (i). This condition can be made even weaker with the help of an additional moment condition. In Assumption 3.4 (ii), we only require that $\mathcal{S}_j$s are combined together to form a set that contains an open ball (in terms of $(\mathbf{x}^{(2)}, \mathbf{p})$). This support requirement is quite mild, and hence it can be satisfied even if each product's characteristic has limited variation across markets. Note also that, if $\operatorname{supp}(\theta_{it})$ is compact, the support of D_{jt} can be compact as well.

It is important to note that we construct $\tilde{\Phi}_j$ without relying on any "thin" (lower-dimensional) subset of the support of the product characteristics as done in the BLP model. Instead, we construct $\tilde{\Phi}_j$ in (3.14) by combining the demand in different markets. This is desirable as estimators that rely on thin or irregular identification may have a slow rate of convergence (Khan and Tamer, 2010). In the pure characteristics model, the individuals have varying tastes (random coefficients) over the product characteristics but not over the products themselves. This is the key feature of the model that allows us to identify the random coefficients through the variation of the product characteristics $(X_t^{(2)}, P_t, D_t)$. In contrast, in the BLP model, there was an additional taste for the product itself, which was the main reason for using the thin set to isolate the demand for each product.

Given Assumption 3.4, we now construct Φ in Eq. (3.1). For each $(\mathbf{x}^{(2)}, \mathbf{p}, \mathbf{d}) \in \bigcup_{j=1}^{J} \mathcal{S}_j$, let $w \equiv (\mathbf{x}^{(2)}, \mathbf{p})/\|(\mathbf{x}^{(2)}, \mathbf{p})\|$ and $u \equiv \mathbf{d}/\|(\mathbf{x}^{(2)}, \mathbf{p})\|$. Define

$$\Phi(w, u) \equiv \tilde{\Phi}_j\Big(\frac{\mathbf{x}^{(2)}}{\|(\mathbf{x}^{(2)}, \mathbf{p})\|}, \frac{\mathbf{p}}{\|(\mathbf{x}^{(2)}, \mathbf{p})\|}, \frac{\mathbf{d}}{\|(\mathbf{x}^{(2)}, \mathbf{p})\|}\Big), \text{ where } (\mathbf{x}^{(2)}, \mathbf{p}, \mathbf{d}) \in \mathcal{S}_j. \tag{3.20}$$

Here, for each $(\mathbf{x}^{(2)}, \mathbf{p}, \mathbf{d})$, any j can be used to construct $\tilde{\Phi}_j$ through marginalization as long as $\mathcal{S}_j$ contains $(\mathbf{x}^{(2)}, \mathbf{p}, \mathbf{d})$. Then Φ is defined on a set that is rich enough to invert the Radon (or limited angle Radon) transform. The rest of the analysis parallels our analysis of the BLP model.[10] We therefore obtain the following point identification result.

Theorem 3.2. *Suppose Assumptions 2.1-3.1, and 3.4 hold. Then, f_θ is identified in the pure characteristics demand model, where $\theta_{it} = (\beta_{it}^{(2)}, \alpha_{it})$.*

4 Extensions

Below, we show that our strategy set forth in the previous section can also be applied to extended models that share the key features of the market level demand model. These extended models involve choices of bundles (Section 4.1) and multiple units of consumption (Section 4.2). For both models, we consider the setting with $\sigma_\epsilon = 1$, but it is also possible to analyze the case without the tastes for products.[11] We further analyze the case in which random coefficients are alternative specific (Section 4.3).

4.1 Bundle choice (Example 2)

We consider an alternative procedure for inverting the demand in Example 2. This is because this example (and also the example in the next section) has 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), Example 2 has 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

[10] Note that the additional independence (or i.i.d.) assumptions on $(\epsilon_{i1t}, \cdots, \epsilon_{iJt})$ is not needed in the pure characteristics model.

[11] For the setting without the tastes for products, we refer to Dunker, Hoderlein, and Kaido (2013), an earlier version of the paper.

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.[12] For this, we assume that the following condition is met.

Condition 4.1. *The random coefficient density f_θ is continuously differentiable.* $(\epsilon_{i1t}, \epsilon_{i2t})$ *and* (D_{1t}, D_{2t}) *have full supports in* $\mathbb{R}^2$ *respectively.*

Let $\tilde{\mathbb{L}} = \{(1,0),(1,1)\}$. From (2.7), it is straightforward to show that $\varphi_{(1,0)}$ is strictly increasing in D_{1t} but is strictly decreasing in D_{2t}, while $\varphi_{(1,1)}$ is strictly increasing both in D_{1t} and D_{2t}. Hence, the Jacobian matrix is non-degenerate. Together with a mild support condition on (D_{1t}, D_{2t}), this allows to invert the demand (sub)system and write $\Xi_{jt} = \psi_j(X_t^{(2)}, P_t, \tilde{S}_t) - X_{jt}^{(1)}$, where $\tilde{S}_t = (S_{(1,0),t}, S_{(1,1),t})$. This ensures Assumption 2.2 in this example (see Lemma B.2 given in the appendix). By Theorem 2.1, one can then nonparametrically identify subcomponents $(\varphi_{(1,0)}, \varphi_{(1,1)})$ of the demand function φ.

One may alternatively choose $\tilde{\mathbb{L}} = \{(0,0),(0,1)\}$, and the argument is similar, which then identifies $(\varphi_{(0,0)}, \varphi_{(0,1)})$, and hence all components of the demand function φ are identified. 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_{1t}, D_{2t}). 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.

Given the demand for bundles, we now analyze identification of the random coefficient density. By (2.5), the demand for bundle (0,0) is given by

$$\begin{aligned}
&\phi_{(0,0)}(x^{(2)}, p, \delta) \\
&\quad = \int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 < -\delta_1\} 1\{x_2^{(2)\prime} b^{(2)} + ap_2 + e_2 < -\delta_2\} \\
&\qquad\qquad \times 1\{(x_1^{(2)} + x_2^{(2)})' b^{(2)} + a(p_1 + p_2) + (e_1 + e_2) + \Delta < -\delta_1 - \delta_2\} f_\theta(b^{(2)}, a, e, \Delta) d\theta.
\end{aligned} \tag{4.1}$$

[12] The additional structure can potentially be tested. 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.

Given product $j \in \{1,2\}$, let $-j$ denote the other product. We then define $\tilde{\Phi}_l$ with $l = (0,0)$ as in the BLP example by letting D_{-jt} take a large negative value. For each $(x^{(2)}, p, \delta)$, let

$$\tilde{\Phi}_{(0,0)}(x_j^{(2)}, p_j, \delta_j) \equiv -\lim_{\delta_{-j}\to-\infty} \phi_{(0,0)}(x^{(2)}, p, \delta), \; j = 1, 2. \tag{4.2}$$

We then define $\Phi_{(0,0)}$ as in (3.5).[13] Consider for the moment $j = 1$ in (4.2). Then, $\Phi_{(0,0)}$ is related to the joint density f_{ϑ_1} of $\vartheta_{i1t} \equiv (\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t})$ through a Radon transform.[14] Arguing as in (3.6), it is straightforward to show that $\partial\Phi_{(0,0)}(w,u)/\partial u = \mathcal{R}[f_{\vartheta_1}](w,u)$ with $w \equiv (x_1^{(2)}, p_1, 1)/\|(x_1^{(2)}, p_1, 1)\|$ and $u \equiv \delta_1/\|(x_1^{(2)}, p_1, 1)\|$. Hence, one may identify f_{ϑ_1} by inverting the Radon transform under Assumptions 3.1 and 3.2 with $J = 2$.

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

The demand for bundle (1,1) is given by

$$\begin{aligned}
&\phi_{(1,1)}(x^{(2)}, p, \delta) \\
&\quad = \int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 + \Delta > -\delta_1\} 1\{x_2^{(2)\prime} b^{(2)} + ap_2 + e_2 + \Delta > -\delta_2\} \\
&\quad \times 1\{(x_1^{(2)} + x_2^{(2)})' b^{(2)} + a(p_1 + p_2) + (e_1 + e_2) + \Delta > -\delta_1 - \delta_2\} f_\theta(b^{(2)}, a, e, \Delta) d\theta.
\end{aligned} \tag{4.3}$$

Note that Δ_{it} can be viewed as an additional random coefficient on the constant whose sign is fixed. Hence, the set of covariates includes a constant. Again, conditioning on an event where D_{-jt} takes a large negative value and normalizing the arguments by the norm of $(x_j^{(2)}, p_j, 1)$ yield a function $\Phi_{(1,1)}$ that is related to the density of $\eta_{ijt} \equiv (\beta_{it}^{(2)}, \alpha_{it}, \Delta_{it} + \epsilon_{ijt})$ through the Radon transform in (3.2). Note that the last component of η_j and ϑ_j differ only in the bundle effect Δ_{it}. Hence, if ϵ_{ijt} is independent of Δ_{it} conditional on $(\beta_{it}^{(2)}, \alpha_{it})$, the distribution of Δ_{it} can be identified via deconvolution. For this, let $\Psi_{\epsilon_j|(\beta^{(2)},\alpha)}$ denote the characteristic function of ϵ_{ijt} conditional on $(\beta_{it}^{(2)}, \alpha_{it})$. We summarize these results below.

Theorem 4.2. *Suppose Assumptions 2.1-3.2, 3.4 and Condition 4.1 hold with $J = 2$ and $\theta_{it} = (\beta_{it}^{(2)}, \alpha_{it}, \Delta_{it}, \epsilon_{i1t}, \epsilon_{i2t})$. Suppose the conditional distribution of ϵ_{ijt} given $(\beta_{it}^{(2)}, \alpha_{it})$ is identical for $j = 1, 2$.*

Then, (a) $f_{\vartheta_j}, f_{\eta_j}$ are nonparametrically identified in Example 2; (b) If, in addition, $\Delta_{it} \perp$

[13] In the BLP example, we invert a Radon transform only once. Hence Φ in (3.5) does not have any subscript. In Examples 2 and 3, we invert Radon transfomrs multiple times, and to make this point clear we add subscripts to Φ (e.g. $\Phi_{(0,0)}$ and $\Phi_{(1,1)}$).

[14] Since the bundle effect Δ_{it} does not appear in (4.1), one may only identify the joint density of the subvector $(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t})$ from the demand for bundle (0,0).

$\epsilon_{ijt}|(\beta_{it}^{(2)}, \alpha_{it})$ and $\Psi_{\epsilon_j|(\beta^{(2)},\alpha)}(t) \neq 0$ for almost all $t \in \mathbb{R}$ and for some j, and $\epsilon_{ijt}, j = 1, 2$ are independently distributed (across j) conditional on $(\beta_{it}^{(2)}, \alpha_{it})$, then f_θ is nonparametrically identified in Example 2.

The identification of the distribution of the bundle effect requires the characteristic function of ϵ_{ijt} to have isolated zeros (see e.g. Devroye, 1989, Carrasco and Florens, 2010). This condition can be satisfied by various distributions including the Type-I extreme value distribution and normal distribution.

Remark 4.1. Note that the conditions of Theorem 4.2 do not impose any sign restriction on Δ_{it}. Hence, the two goods can be substitutes ($\Delta_{it} < 0$) for some individuals and complements ($\Delta_{it} > 0$) for others. This feature, therefore, can be useful for analyzing bundles of goods whose substitution pattern can significantly differ across individuals (e.g. E-books and print books).

Remark 4.2. We note that the utility specification adopted in the pure characteristics model can also be combined with the bundle choice and multiple units of consumption studied next. The identification of the random coefficients can be achieved using arguments similar to the ones in Section 3.2[15].

4.2 Multiple units of consumption (Example 3)

One may also consider settings where multiple units of consumption are allowed. For simplicity, we consider the simplest setup where $J = 2$ and $Y_1 \in \{0, 1, 2\}$ 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:

$$U^*_{i,(y_1,y_2),t} = y_1 U^*_{i1t} + y_2 U^*_{i2t} + \Delta_{i,(y_1,y_2),t} \ , \tag{4.4}$$

where $\Delta_{i,(y_1,y_2),t}$ 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^*_{i1t} and U^*_{i2t} give the utility from consuming a single unit of each of the two goods. Throughout this example, we assume that $U^*_{i,(y_1,y_2),t}$ is concave in (y_1, y_2). Then, a bundle is chosen if its utility exceeds those of the neighboring alternatives. For example, bundle $(2, 0)$ is chosen if it is preferred to

[15] The analysis of these settings are contained in an earlier version of this paper, which is available from the authors upon request.

bundles (1,0), (1,1) and (2,1). That is,

$$
\begin{aligned}
&2(X_{1t}'\beta_{it} + \alpha_{it}P_{1t} + \Xi_{1t} + \epsilon_{i1t}) + \Delta_{i,(2,0),t} > X_{1t}'\beta_{it} + \alpha_{it}P_{1t} + \Xi_{1t} + \epsilon_{i1t} \ , \\
&2(X_{1t}'\beta_{it} + \alpha_{it}P_{1t} + \Xi_{1t} + \epsilon_{i1t}) + \Delta_{i,(2,0),t} \\
&\qquad\qquad > X_{1t}'\beta_{it} + \alpha_{it}P_{1t} + \Xi_{1t} + \epsilon_{i1t} + X_{2t}'\beta_{it} + \alpha_{it}P_{2t} + \Xi_{2t} + \epsilon_{i2t} + \Delta_{i,(1,1),t} \\
&2(X_{1t}'\beta_{it} + \alpha_{it}P_{1t} + \Xi_{1t} + \epsilon_{i1t}) + \Delta_{i,(2,0),t} \ , \\
&\qquad\qquad > 2(X_{1t}'\beta_{it} + \alpha_{it}P_{1t} + \Xi_{1t} + \epsilon_{i1t}) + X_{2t}'\beta_{it} + \alpha_{it}P_{2t} + \Xi_{2t} + \epsilon_{i2t} + \Delta_{i,(2,1),t}. \qquad (4.5)
\end{aligned}
$$

The aggregate structural demand can be obtained as

$$
\begin{aligned}
\varphi_{(2,0)}(X_t, P_t, \Xi_t) = \int & 1\{X_{1t}'b + aP_{1t} + e_1 + \Delta_{(2,0)} > -\Xi_{1t}\} \\
&\times 1\{(X_{1t} - X_{2t})'b + a(P_{1t} - P_{2t}) + (e_1 - e_2) + \Delta_{(2,0)} - \Delta_{(1,1)} > -\Xi_{1t} + \Xi_{2t}\} \\
&\qquad \times 1\{X_{2t}'b + aP_{2t} + e_2 + \Delta_{(2,1)} - \Delta_{(2,0)} < -\Xi_{2t}\} f_\theta(b, a, e, \Delta) d\theta \ . \quad (4.6)
\end{aligned}
$$

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

Let $\tilde{\mathbb{L}} = \{(2,0), (2,1)\}$. From (4.5), $\varphi_{(2,0)}$ 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. This ensures Assumption 2.2 in this example, and by Theorem 2.1, one can then nonparametrically identify subcomponents $\{\varphi_l, l \in \tilde{\mathbb{L}}\}$ of the demand function φ. 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 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: $\theta_{it} = (\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t}, \epsilon_{i2t}, \Delta_{i,(1,1),t}, \Delta_{i,(2,0),t}, \Delta_{i,(2,1),t})'$. We begin with the demand for $(0,0)$, $(0,1)$, $(2,0)$, and $(2,1)$ given by

$$\begin{aligned}
\phi_{(0,0)}(x^{(2)},p,\delta) = \int & 1\{x_1^{(2)\prime}b^{(2)} + ap_1 + e_1 < -\delta_1\} \\
& \times 1\{x_2^{(2)\prime}b^{(2)} + ap_2 + e_2 < -\delta_2\} \\
& \times 1\{(x_1^{(2)} + x_2^{(2)})'b^{(2)} + a(p_1+p_2) + (e_1+e_2) < -\delta_1 - \delta_2\} f_\theta(b^{(2)},a,e,\Delta) d\theta \ , \\
\phi_{(0,1)}(x^{(2)},p,\delta) = \int & 1\{x_2^{(2)\prime}b^{(2)} + ap_2 + e_2 > -\delta_2\} \\
& \times 1\{(x_1^{(2)} - x_2^{(2)})'b^{(2)} + a(p_1-p_2) + (e_1-e_2) < -\delta_1 + \delta_2\} \\
& \times 1\{x_1^{(2)\prime}b^{(2)} + ap_1 + e_1 + \Delta_{(1,1)} > -\delta_1\} f_\theta(b^{(2)},a,e,\Delta) d\theta \ , \\
\phi_{(2,0)}(x^{(2)},p,\delta) = \int & 1\{x_1^{(2)\prime}b^{(2)} + ap_1 + e_1 + \Delta_{(2,0)} > -\delta_1\} \\
& \times 1\{(x_1^{(2)} - x_2^{(2)})'b^{(2)} + a(p_1-p_2) + (e_1-e_2) + \Delta_{(2,0)} - \Delta_{(1,1)} > -\delta_1 + \delta_2\} \\
& \times 1\{x_2^{(2)\prime}b^{(2)} + ap_2 + e_2 + \Delta_{(2,1)} - \Delta_{(2,0)} < -\delta_2\} f_\theta(b^{(2)},a,e,\Delta) d\theta \ , \\
\phi_{(2,1)}(x^{(2)},p,\delta) = \int & 1\{x_1^{(2)\prime}b^{(2)} + ap_1 + e_1 + \Delta_{(2,1)} - \Delta_{(1,1)} > -\delta_1\} \\
& \times 1\{x_1^{(2)\prime}b^{(2)} + ap_1 + e_1 + \Delta_{(2,1)} - \Delta_{(2,0)} > -\delta_2\} \\
& \times 1\{(x_1^{(2)} + x_2^{(2)})'b^{(2)} + a(p_1+p_2) + (e_1+e_2) + \Delta_{(2,1)} > -\delta_1 - \delta_2\} f_\theta(b^{(2)},a,e,\Delta) d\theta \ .
\end{aligned}$$

Hence, if D_{2t} has a large support, by taking δ_2 sufficiently small or sufficiently large, we may define

$$\begin{aligned}
\tilde{\Phi}_{(0,0)}(x_1^{(2)},p_1,\delta_1) &\equiv -\lim_{\delta_2\to-\infty} \phi_{(0,0)}(x^{(2)},p,\delta) \\
&= -\int 1\{x_1^{(2)\prime}b^{(2)} + ap_1 + e_1 < -\delta_1\} f_\theta(b^{(2)},a,e,\Delta) d\theta \ , && (4.7) \\
\tilde{\Phi}_{(0,1)}(x_1^{(2)},p_1,\delta_1) &\equiv -\lim_{\delta_2\to\infty} \phi_{(0,1)}(x^{(2)},p,\delta) \\
&= -\int 1\{x_1^{(2)\prime}b^{(2)} + ap_1 + e_1 + \Delta_{(1,1)} > -\delta_1\} f_\theta(b^{(2)},a,e,\Delta) d\theta \ , && (4.8) \\
\tilde{\Phi}_{(2,0)}(x_1^{(2)},p_1,\delta_1) &\equiv -\lim_{\delta_2\to-\infty} \phi_{(2,0)}(x^{(2)},p,\delta) \\
&= -\int 1\{x_1^{(2)\prime}b^{(2)} + ap_1 + e_1 + \Delta_{(2,0)} > -\delta_1\} f_\theta(b^{(2)},a,e,\Delta) d\theta \ , && (4.9) \\
\tilde{\Phi}_{(2,1)}(x_1^{(2)},p_1,\delta_1) &\equiv -\lim_{\delta_2\to\infty} \phi_{(2,1)}(x^{(2)},p,\delta) \\
&= \int 1\{x_1^{(2)\prime}b^{(2)} + ap_1 + e_1 + \Delta_{(2,1)} - \Delta_{(1,1)} > -\delta_1\} f_\theta(b^{(2)},a,e,\Delta) d\theta \ . && (4.10)
\end{aligned}$$

For each $l \in \{(0,0),(0,1),(2,0),(2,1)\}$, define Φ_l as in (3.5). Arguing as in Example 2, Φ_l is then related to the random coefficient densities by

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

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, for each l, f_{ϑ_l} is the joint density of a subvector $\vartheta_{i,l,t}$ of θ_{it}, which is given by[16]

$$\vartheta_{i,(0,0),t} = (\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t}),\ \vartheta_{i,(0,1),t} = (\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t} + \Delta_{i,(1,1),t}),$$
$$\vartheta_{i,(2,0),t} = (\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t} + \Delta_{i,(2,0),t}),\ \vartheta_{i,(2,1),t} = (\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t} + \Delta_{i,(2,1),t} - \Delta_{i,(1,1),t}). \quad (4.11)$$

The joint density of θ_{it} is identified by making the following assumption.

Assumption 4.1. *(i)* $(\Delta_{i,(1,1),t}, \Delta_{i,(2,0),t}, \Delta_{i,(2,1),t}) \perp \epsilon_{ijt} | (\beta_{it}^{(2)}, \alpha_{it})$ *and* $\Psi_{\epsilon_j|(\beta^{(2)},\alpha)}(t) \neq 0$ *for almost all* $t \in \mathbb{R}$ *and for some* $j \in \{1,2\}$*; (ii)* $\epsilon_{ijt}, j = 1,2$ *are independently and identically distributed (across j) conditional on* $(\beta_{it}^{(2)}, \alpha_{it})$*; (iii)* $(\Delta_{i,(1,1),t}, \Delta_{i,(2,0),t}, \Delta_{i,(2,1),t})$ *are independent of each other conditional on* $(\beta_{it}^{(2)}, \alpha_{it})$ *and* $\Psi_{\Delta_{(1,1)}|(\beta^{(2)},\alpha)}(t) \neq 0$ *for almost all* $t \in \mathbb{R}$*.*

Assumption 4.1 (iii) 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 4.1 (iii) also adds a regularity condition for recovering the distribution of $\Delta_{i,(2,1),t}$ from those of $\Delta_{i,(2,1),t} - \Delta_{i,(1,1),t}$ and $\Delta_{i,(1,1),t}$ 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 the nonparametric IV step. To see this, we note that the demand for this bundle is given by

$$\phi_{(1,0)}(x^{(2)}, p, \delta) = \int 1\{0 < x_1^{(2)\prime(2)} + ap_1 + e_1 + \delta_1 < -\Delta_{(2,0)}\}$$
$$\times 1\{x_2^{(2)\prime(2)} + ap_2 + e_2 + \delta_2 < -\Delta_{(1,1)}\} 1\{(x_1^{(2)} - x_2^{(2)})^{\prime(2)} + a(p_1 - p_2) + (e_1 - e_2) < -(\delta_1 - \delta_2)\}$$
$$\times 1\{(x_1^{(2)} + x_2^{(2)})' b^{(2)} + a(p_1 + p_2) + (e_1 + e_2) + \Delta_{(2,1)} < -(\delta_1 + \delta_2)\} f_\theta(b^{(2)}, a, e, \Delta) d\theta. \quad (4.12)$$

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.[17]

[16] 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_{1t} would allow one to recover the joint density of $(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i2t} + \Delta_{i,(2,1),t} - \Delta_{i,(2,0),t})$ from the demand for bundle (2,0).

[17] For simplicity, we only consider the case where $\delta_2 \to -\infty$ or ∞ in (4.7)-(4.8). This requires a full support condition on D_{1t}. It is possible to replace this assumption with an analog of Assumption 3.3 by also considering the case where $\delta_1 \to -\infty$ or ∞ and imposing an additional restriction on the distribution of $(\epsilon_{i1t}, \epsilon_{i2t}, \Delta_{i,(1,1),t}, \Delta_{i,(2,0),t}, \Delta_{i,(2,1),t})$.

Theorem 4.3. *Suppose $U_{(y_1,y_2),t}$ is concave in (y_1, y_2). Suppose Condition 4.1 and Assumptions 2.1, 2.3-3.1, 3.4 hold with $J = 2$ and $\theta_{it} = (\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t}, \epsilon_{i2t}, \Delta_{i,(1,1),t}, \Delta_{i,(2,0),t}, \Delta_{i,(2,1),t})$. Suppose that (X_{1t}, P_{1t}, D_{1t}) has a full support. Then, (a) $f_{\vartheta_l}, l \in \{(0,0), (0,1), (2,0), (2,1)\}$ are nonparametrically identified in Example 3; (b) Suppose further that Assumption 4.1 holds. Then, f_θ is identified in Example 3. Further, all components of the structural demand ϕ are identified.*

4.3 Alternative specific coefficients

So far, we have maintained the assumption that $(\beta_{ijt}, \alpha_{ijt}) = (\beta_{it}, \alpha_{it}), \forall j$ almost surely. This excludes alternative specific random coefficients. However, this is not essential in our analysis. One may allow some or all components of $(\beta_{ijt}, \alpha_{ijt})$ to be different random variables across j and identify their joint distribution under an extended support condition on the product characteristics.

We first note that the aggregate demand is identified as long as Assumptions 2.1-2.3 hold. In the BLP model, the marginal density f_{ϑ_j} of $\vartheta_{ijt} = (\beta_{ijt}^{(2)}, \alpha_{ijt}, \epsilon_{ijt})$ can be identified for any j as long as the corresponding product characteristics $(X_{jt}^{(2)}, P_{jt}, D_{jt})$ has a full support using the same identification strategy in Section 3 (see Remark 3.2). For the pure characteristics demand model, we note that the maps $\mathcal{M}_{\mathcal{J}}$ cannot be used as the use of this map is justified when $(\beta_{ijt}, \alpha_{ijt}) = (\beta_{it}, \alpha_{it}), \forall j$. However, the large support assumption $\text{supp}\,(D_{kt}) = \mathbb{R}$ for $k \neq j$ can still be used to construct Φ. Hence, the analysis of this case becomes similar to the BLP model. In both models, the joint density f_θ of $\theta_{it} = (\vartheta_{i1t}, \cdots, \vartheta_{iJt})$ can be recovered under the assumption that ϑ_{ijt} are independent across j.

When the covariates $(X_t^{(2)}, P_t, D_t)$ have rich variations jointly, it is also possible to identify the joint density f_θ without the independence assumption invoked above. This requires us to extend our identification strategy. To see this, we take Example 2 as an illustration below. Consider identifying the joint density of $\theta_{it} = (\beta_{i1t}^{(2)}, \beta_{i2t}^{(2)}, \alpha_{i1t}, \alpha_{i2t}, \epsilon_{i1t}, \epsilon_{i2t} + \Delta_{it})$ under the assumption that the two goods are complements, i.e. $\Delta_{it} > 0, a.s.$ In this setting, we may use the demand for bundle $(1, 0)$, which can be written as

$$\begin{aligned}\phi_{(1,0)}(x^{(2)}, p, \delta) = \int & 1\{x_1^{(2)\prime} b_1^{(2)} + a_1 p_1 + e_1 > -\delta_1\} \\ & \times 1\{x_2^{(2)\prime} b_2^{(2)} + a_2 p_2 + e_2 + \Delta < -\delta_2\} f_\theta(b_1^{(2)}, b_2^{(2)}, a_1, a_2, \Delta) d\theta. \quad (4.13)\end{aligned}$$

To recover the joint density, one has to directly work with this demand function without simplifying it further. A key feature of (4.13) 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 (4.13) involves $(\beta_{i1t}^{(2)}, \alpha_{i1t}, \epsilon_{i1t})$, while the second indicator function involves $(\beta_{i2t}^{(2)}, \alpha_{i2t}, \epsilon_{i2t} + \Delta_{it})$. 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

$$\frac{\partial^2 \phi_{(1,0)}(w_1, w_2, u_1, u_2)}{\partial u_1 \partial u_2} = (\mathcal{R} \otimes \mathcal{R})[f_\theta](w_1, w_2, u_1, -u_2), \tag{4.14}$$

where $w_1 = -(x_1^{(2)}, p_1, 1)/\|(x_1^{(2)}, p_1, 1)\|$, $w_2 = (x_2^{(2)}, p_2, 1)/\|(x_2^{(2)}, p_2, 1)\|$, $u_1 = -\delta_1/\|(x_1^{(2)}, p_1, 1)\|$, $u_2 = \delta_2/\|(x_2^{(2)}, p_2, 1)\|$, 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 (4.14) to identify f_θ requires Assumption 3.3 (i) to be strengthened as follows.

Assumption 4.2. $(X_{1t}^{(2)}, P_{1t}, D_{1t}, X_{2t}^{(2)}, P_{2t}, D_{2t})$ *has a full support.*

This is a stronger support condition than Assumption 3.3 (i) as it requires a joint full support condition 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.[18]

Theorem 4.4. *In Example 2, let* $\theta_{it} = (\beta_{i1t}^{(2)}, \beta_{i2t}^{(2)}, \alpha_{i1t}, \alpha_{i2t}, \epsilon_{i1t}, \epsilon_{i2t} + \Delta_{it})$. *Suppose that Assumptions 2.1-2.3, 3.1, and 4.2 hold. Suppose further that* $\Delta_{it} > 0$, *a.s. Then,* f_θ *is identified.*

5 Suggested estimation methods

5.1 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 from the estimated $\widehat{\psi}_j$. The last step of the estimation is the inversion of a Radon transform.

[18] We omit the proof of this result for brevity. Similar to Theorem 3.1, it is also possible to establish identification using an analog of Assumption 3.3 (ii), which relaxes the support requirement at the cost of an additional moment condition. We also note that one may disentangle the distribution of Δ_{it} from that of $\epsilon_{i2t} + \Delta_{it}$ using a deconvolution argument as done in Theorem 4.2.

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, and regularized inversion schemes must be applied. We propose Tikhonov regularization for this purpose:

$$\widehat{\psi}_j := \min_{\psi} \|\widehat{T}\psi - x_{jt}^{(1)}\|^2_{L^2(X_t, Z_t)} + \alpha \mathfrak{R}(\psi). \tag{5.1}$$

Here, $\alpha \geq 0$ is a regularization parameter and $\mathfrak{R}$ a regularization functional. A common choice is $\mathfrak{R}(\psi) = \|\psi\|^2_{L^2}$, however, if more smoothness is expected, this can be a squared Sobolev norm or some other norm. 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. Since monotonicity is a convex constraint, even with this choice of $\mathfrak{R}$, equation (5.1) is a convex minimization problem. Solving the problem is computationally feasible, see 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.

In the second step $\widehat{\psi}_j(X_t^{(2)}, P_t, S_t)$ is inverted in S_t to get an estimate $\widehat{\phi}_j$ for the demand function ϕ_j. In the BLP model, we approximate the limit of $\widehat{\psi}_j(X_t^{(2)}, P_t, S_t)$ for $D_{kt} \to -\infty$ to construct an estimate for $\tilde{\Phi}_j$ as in (3.4). When ϵ_{ijt} is iid across j, one may improve efficiency by repeating this process for all products and averaging $\tilde{\Phi}_j$ across $j = 1, \cdots, J$. For the pure characteristics model an estimate of $\tilde{\Phi}_j$ is computed from $\widehat{\phi}_j$ by a sum over permutations as in (3.14). Similar constructions can be carried out for the models of bundel choices (4.2) and multiple unites of consumption (4.7) – (4.10). From an estimator of $\tilde{\Phi}_j$ we get an estimate $\widehat{\Phi}$ of Φ by normalization as in (3.5) or (3.20).

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 $\vartheta = (b, a, e)$ in the BLP model, $\vartheta = (b, a)$ in the PCM, or $\vartheta = (b, a, \Delta)$ in other models. The operator $(R^* g)(x) := \int_{\|w\|=1} g(w, w'x) dw$ 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.

5.2 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 $\theta_{it} = (\beta_{it}^{(2)}, \alpha_{it}, \Delta_{it}, \epsilon_{1it}, \epsilon_{2it})$ 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_{1t}, D_{2t}), where $D_{jt} = \Xi_{jt} + X_{jt}^{(1)}, j = 1, 2$ and obtain the inversion in eq. (2.8).

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. 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, if the conditional CDF of ϵ_{ijt} given $(\beta_{it}^{(2)}, \alpha_{it}, \Delta_{it})$ has an analytic form, the two-step method in BLP and Berry and Pakes (2007) can be employed. We take the demand for bundle (0,0) in eq. (4.1) as an example. Conditional on the product characteristics $y \equiv (x^{(2)}, p, \delta)$ and the rest of the random coefficients $(\beta_{it}^{(2)}, \alpha_{it}, \Delta_{it})$, bundle (0,0) is chosen when

$$\sigma_\epsilon \epsilon_{i1t} < h_1(y, b^{(2)}, a, \Delta) \quad \text{and} \quad \sigma_\epsilon \epsilon_{i2t} < h_2(y, b^{(2)}, a, \Delta), \quad \text{if} \;\; \Delta < 0 \tag{5.2}$$

$$\sigma_\epsilon \epsilon_{i1t} < h_2(y, b^{(2)}, a, \Delta) \quad \text{and} \quad \sigma_\epsilon (\epsilon_{i1t} + \epsilon_{i2t}) < h_3(y, b^{(2)}, a, \Delta), \quad \text{if} \;\; \Delta \geq 0, \tag{5.3}$$

where

$$h_1(y,\beta^{(2)},a,\Delta) \equiv -x_1^{(2)\prime}b^{(2)} - ap_1 - \delta_1, \quad h_2(y,\beta^{(2)},a,\Delta) \equiv -x_2^{(2)\prime}b^{(2)} - ap_2 - \delta_2,$$
$$h_3(y,\beta^{(2)},a,\Delta) \equiv -(x_1^{(2)} + x_2^{(2)})'^{(2)} - a(p_1+p_2) - (\delta_1 - \delta_2). \quad (5.4)$$

In what follows, we consider the BLP setting where $\sigma_\epsilon = 1$.[19] Specify the conditional distribution of $(\epsilon_{i1t}, \epsilon_{i2t})$ given $(\beta_{it}^{(2)}, \alpha_{it}, \Delta_{it})$. For each $(y, b^{(2)}, a, \Delta)$, define

$$G(y,b^{(2)},a,\Delta) \equiv \begin{cases} Pr(\epsilon_{i1t} < h_1(y,b^{(2)},a,\Delta),\ \epsilon_{i2t} < h_2(y,b^{(2)},a,\Delta)|y,b^{(2)},a,\Delta) & \Delta < 0 \\ Pr(\epsilon_{i1t} < h_2(y,b^{(2)},a,\Delta),\ \epsilon_{i1t}+\epsilon_{i2t} < h_3(y,b^{(2)},a,\Delta)|y,b^{(2)},a,\Delta) & \Delta > 0. \end{cases} \quad (5.7)$$

The value of $G(y, b^{(2)}, a, \Delta)$ can be calculated analytically, for example, if one specifies the joint distribution of $(\epsilon_{i1t}, \epsilon_{i2t})$ as normal. Eq. (5.2)-(5.3) then imply that the aggregate share of bundle (0,0) is given by

$$\phi_{(0,0)}(x^{(2)}, p, \delta; \gamma) = \int G(y, b^{(2)}, a, \Delta) f_{\beta^{(2)},a,\Delta}(b, a, \Delta; \gamma) d\theta. \quad (5.8)$$

This can be approximated by the simulated moment:

$$\hat{\phi}_{(0,0)}(x^{(2)}, p, \delta; \gamma) = \frac{1}{n_S} \sum_{i=1}^{n_S} G(y, b_i^{(2)}, a_i, \Delta_i), \quad (5.9)$$

where the simulated sample $\{(b_i^{(2)}, a_i, \Delta_i), i = 1, \cdots, n_S\}$ is generated from $f_{\beta^{(2)},a,\Delta}(\cdot;\gamma)$.[20] 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 $\tilde{\mathbb{L}} = \{(0,0),(0,1)\}$. Then, $(\delta_1, \delta_2) \mapsto (\hat{\phi}_{(0,0)}(x^{(2)}, p, \delta; \gamma), \hat{\phi}_{(0,1)}(x^{(2)}, p, \delta; \gamma))$ 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

[19] In the PCM, one may adopt a similar approach by letting one of the remaining random coefficients play the role of ϵ_{ijt}. For example, replace (5.2)-(5.3) with

$$a < h_1(y, b^{(2)}, \Delta), \text{ and } a < h_2(y, b^{(2)}, \Delta), \text{ if } \Delta < 0 \quad (5.5)$$
$$a < h_2(y, b^{(2)}, \Delta), \text{ and } a < h_3(y, b^{(2)}, \Delta), \text{ if } \Delta \geq 0, \quad (5.6)$$

where $h_1(y, b^{(2)}, \Delta) = (-x_1^{(2)\prime(2)} - \delta_1)/p_1$, and h_2, h_3 are defined similarly. Specify the conditional distribution of α_{it} so that an analog of (5.7) can be calculated. The rest of the estimation procedure is similar.

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

the inverse of this mapping[21], which then yields $\hat{\psi}(\cdot;\gamma) \equiv (\hat{\psi}_1(\cdot;\gamma), \hat{\psi}_2(\cdot;\gamma))$ such that

$$\Xi_{1,t} = \hat{\psi}_1(X_t^{(2)}, P_t, S_{(0,0),t}, S_{(0,1),t};\gamma) - X_{1t}^{(1)}, \quad \Xi_{2,t} = \hat{\psi}_2(X_t^{(2)}, P_t, S_{(0,0),t}, S_{(0,1),t};\gamma) - X_{2t}^{(1)} \tag{5.10}$$

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

$$\Xi_{1,t} = \hat{\psi}_3(X_t^{(2)}, P_t, S_{(1,0),t}, S_{(1,1),t};\gamma) - X_{1t}^{(1)}, \quad \Xi_{2,t} = \hat{\psi}_4(X_t^{(2)}, P_t, S_{(1,0),t}, S_{(1,1),t};\gamma) - X_{2t}^{(1)} \tag{5.11}$$

This helps generate additional moment restrictions in the next step.

The third step is to use (5.10)-(5.11) 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:

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

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.

6 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

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

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., A. Feuerverger, and P. Hall (1996): "On nonparametric estimation of intercept and slope distributions in random coefficient regression," *The Annals of Statistics*, 24(6), 2569–2592.

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

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

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

[9] ——— (2013): "Identification in Differentiated Products Markets Using Market Level Data," Cowles Foundation Discussion Paper.

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

[11] ——— (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.

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

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

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

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

[16] 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.

[17] Dunker, F. (2015): "Convergence of the risk for nonparametric IV quantile regression and nonparametric IV regression with full independence," *CRC-PEG Discussion Paper* , No. 192.

[18] 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.

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

[20] Dunker, F., S. Hoderlein, and H. Kaido (2014): "Nonparametric Identification of Endogenous and Heterogeneous Aggregate demand models: Complements, Bundles and the Market Level," CEMMAP Working Paper.

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

[22] Fox, J., K. Kim, S. Ryan, and P. Bajari (2012): "The random coefficients logit model is identified," *Journal of Econometrics*, 106, 204–212.

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

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

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

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

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

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

[30] Gowrisankaran, G and M. Rysman (2012): "Dynamics of Consumer Demand for New Durable Goods," *Journal of Political Economy*, 120, 1173-1219.

[30] Heckman, J. J., R. L. Matzkin and L. Nesheim (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78(5), 1569–1591.

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

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

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

[34] 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.

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

[36] Khan, S., and E. Tamer (2010) "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 78(6), 2021–2042.

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

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

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

[40] 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.

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

[42] 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.

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

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

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

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

A Notation and Definitions

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

$\mathbb{S}^{q-1}$: The unit sphere $\mathbb{S}^{q-1} \equiv \{v \in \mathbb{R}^q : \|v\| = 1\}$.
$\mathbb{H}_+$: The hemisphere $\mathbb{H}_+ \equiv \{v = (v_1, v_2, \cdots, v_q) \in \mathbb{S}^{q-1} : v_q \geq 0\}$.
$P_{w,r}$: The hyperplane: $P_{w,r} \equiv \{v \in \mathbb{R}^q : v'w = r\}$.
$\mu_{w,r}$: Lebesgue measure on $P_{w,r}$.
$\mathcal{R}$: Radon transform: $\mathcal{R}[f](w,u) = \int_{P_{w,u}} f(v) d\mu_{w,u}(v)$.

B Proofs

Proof of Theorem 2.1. 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). □

Lemma B.1. *Let $\theta = (\theta_1, \ldots; \theta_d)$ be a d-dimensional random vector with density f_θ. Assume that the moments of all components are finite $\mathbb{E}[|\theta_d|^l] < \infty$ for all $i = 1, \ldots, d$ and $l = \mathbb{N}$. In addition, let for any $z > 0$*

$$0 = \lim_{p \to \infty} \frac{z^l}{l!} E\left[(|\theta_1| + |\theta_2| + \ldots + |\theta_d|)^l \right] .$$

For any open neighborhood $\mathcal{U} \subset \mathbb{S}^{d-1}$ it holds that if the Radon transform of $\mathcal{R}[f_\theta](w, \delta)$ is known for all $(w, \delta) \in \{(w, w't) | w \in \mathcal{U}, t \in \operatorname{supp}(\theta)\}$, the density f_θ is identified.

Proof of Lemma B.1. We first show that $\mathcal{F}f_\theta$ the Fourier transform of f_θ is analytic. The Fourier transform can be approximated by the p-th Taylor polynomial for some point $b_0 \in \mathbb{R}^d$. The Taylor remainder for some point $b \in \mathbb{R}^d$ is bounded by

$$R_p(\mathcal{F}f_\theta)(b; b_0) \leq \sum_{\alpha \in \mathbb{N}^d, |\alpha| = p+1} \frac{(b - b_0)^\alpha}{\alpha!} \|D^\alpha \mathcal{F} f_\theta\|_\infty .$$

In this formula the multi-index notation is used with respect to α. This means $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d) \in$

$\mathbb{N}^d$, $|\alpha| := \sum_{i=1}^d \alpha_i$, $\alpha! := \prod_{i=1}^d \alpha_i!$, and

$$D^\alpha \mathcal{F} f_\theta = \frac{\partial^{|\alpha|} \mathcal{F} f_\theta}{\partial b_1^{\alpha_1} \partial b_2^{\alpha_2} \dots \partial b_k^{\alpha_k}}.$$

Note that

$$\begin{aligned}
\|D^\alpha \mathcal{F} f_\theta\|_\infty &\leq \int_{\mathbb{R}^d} |v_1^{\alpha_1} v_2^{\alpha_2} \dots v_d^{\alpha_d}| f_\theta(v_1, v_2, \dots, v_d) dv \\
&\leq \int_{\mathbb{R}^d} |v_1|^{\alpha_1} |v_2|^{\alpha_2} \dots |v_d|^{\alpha_d} f_\theta(v_1, v_2, \dots, v_d) dv \\
&= E[|\theta_2|^{\alpha_1} |\theta_2|^{\alpha_2} \dots |\theta_d|^{\alpha_k}|].
\end{aligned}$$

This yields

$$\begin{aligned}
R_p(\mathcal{F} f_\theta)(b; b_0) &\leq \|b - b_0\|_\infty^p E\left[\sum_{\alpha \in \mathbb{N}^d, |\alpha| = p+1} \frac{|\theta_1|^{\alpha_1} |\theta_2|^{\alpha_2} \dots |\theta_d|^{\alpha_d}}{\alpha!}\right] \\
&\leq \|b - b_0\|_\infty^p E\left[(p!)^{-1} \left(|\theta_1| + |\theta_2| + \dots + |\theta_d|\right)^p\right] \\
&\leq \frac{\|b - b_0\|_\infty^p}{p!} E\left[\left(|\theta_1| + |\theta_2| + \dots + |\theta_d|\right)^p\right].
\end{aligned}$$

Hence, the Taylor approximation converges point-wise to $\mathcal{F} f_\theta$ on $\mathbb{R}^d$. Consequently, if $\mathcal{F} f_\theta$ is know on some neighborhood around b_0, $\mathcal{F} f_\theta$ is identified. This makes $\mathcal{F} f_\theta$ an analytic function. Since the Fourier transform is bijectiv this identifies f_θ as well.

It remains to show that $\mathcal{F} f_\theta$ is known in some open neighborhood. By the Fourier slice theorem for the Radon transform $(\mathcal{F} f_\theta)(w\eta) = \mathcal{F}_1(\mathcal{R} f_\theta[w, \cdot])(\eta)$. Here $\mathcal{F}_1$ denotes the one-dimensional Fourier transform that acts on the free variable denoted by " $\cdot$ ". Note that $\mathcal{R} f_\theta[w, \delta] = 0$ if $w \in \mathcal{U}$ but $(w, \delta) \notin \left\{(w, w't) | w \in \mathcal{U}, t \in \operatorname{supp}(\theta)\right\}$. Thus, if $\mathcal{R} f_\theta[w, \delta]$ is known for all $(w, \delta) \in \left\{(w, w't) | w \in \mathcal{U}, t \in \operatorname{supp}(\theta)\right\}$, it is known for all $w \in \mathcal{U}$ and all $\delta \in \mathbb{R}$. It follows that $\mathcal{F} f_\theta$ is known on some open neighborhood. This identifies f_θ. □

Proof of Theorem 3.1. (i) 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 2.1, ψ is identified. Further, the aggregate demand ϕ is identified by (2.10) and the identity $\phi_0 = 1 - \sum_{j=1}^J \phi_j$.

For any product j and product characteristics $(x_j^{(2)}, p_j, \delta_j)$ define the new function

$$\tilde{\Phi}_j(x_j^{(2)}, p_j, \delta_j) = - \lim_{\delta_1, \dots, \delta_{j-1}, \delta_{j+1}, \dots, \delta_J \to -\infty} \phi_j(x^{(2)}, p, \delta)$$

point wise. Here $\phi_j(x^{(2)}, p, \delta)$ can be any fixed vector of product characteristics where $(x_j^{(2)}, p_j, \delta_j)$ coincide with the values on the l.h.s. of the equation. The limit on the r.h.s. exists and is unique. This can be seen by using the definition of ϕ_j, Lebesgue's theorem, and Assumption 3.2. Consequently,

$$\tilde{\Phi}(x_j^{(2)}, p_j, \delta_j) = -\int 1\{x_j^{(2)\prime} b^{(2)} + ap_j + \epsilon_j < -\delta_j\} f_{\vartheta_j}(b^{(2)}, a, e_j) d\vartheta_j.$$

Now define Φ as in (3.5) and conclude

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

Taking a derivative with respect to u yields (3.7). By the assumption that the conditional distribution of ϵ_{ijt} given $(\beta_{it}^{(2)}, \alpha_{it})$ is identical for $j = 1, \cdots, J$, it follows that $f_{\vartheta_j} = f_\vartheta, \forall j$ for some common density f_ϑ. Hence, we may rewrite (3.7) as

$$\frac{\partial \Phi(w, u)}{\partial u} = \mathcal{R}[f_\vartheta](w, u). \tag{B.2}$$

Note that by Assumptions 3.1 (i) and 3.2, $\partial\Phi(w, u)/\partial u$ is well-defined for some $(w, u) \in \mathbb{H}_+ \times \mathbb{R}$. By Assumption 3.3 $\partial\Phi(w, u)/\partial u$ is either identified for all $(w, u) \in \mathbb{H}_+ \times \mathbb{R}$ or only for w in some open neighborhood of $\mathbb{H}_+$. In the first case the identification of f_ϑ follows from Assumption from the injectivity of the Radon transform (Theorem I in Cramér and Wold, 1936). In the second case the the identification of f_ϑ follows from Lemma B.1.

(ii) In the first part of the proof f_{ϑ_j}, $j = 1, 2, \ldots, J$ were identified (as f_ϑ). Hence, the conditional distribution $f_{\epsilon_j|\beta^{(2)},\alpha}$ of ϵ_{ijt} given $(\beta_{it}^{(2)}, \alpha_{it})$ and the marginal distribution $f_{\beta^{(2)},\alpha}$ of $(\beta_{it}^{(2)}, \alpha_{it})$ are identified for any j. Under the additional assumption that $\epsilon_{i1t}, \epsilon_{i2t}, \ldots, \epsilon_{iJt}$ are independent conditional on $(\beta_{it}^{(2)}, \alpha_{it})$ we get the joint distribution of θ_{it} by

$$f_\theta(b^{(2)}, \alpha, e_1, \ldots, e_J) = \prod_{j=1}^{J} f_{\epsilon_j|\beta^{(2)},\alpha}(e_j | b^{(2)}, \alpha) \times f_{\beta^{(2)},\alpha}(b^{(2)}, \alpha). \tag{B.3}$$

Hence, f_θ is identified. □

Proof of Theorem 3.2. 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 2.1, ψ is identified. Further, the aggregate demand ϕ is identified by (2.10) and the identity $\phi_0 = 1 - \sum_{j=1}^{J} \phi_j$. By Assumption 3.4, for each $(\mathbf{x}^{(2)}, \mathbf{p}, \mathbf{d}) \in \mathbb{R}^{d_X - 1} \times \mathbb{R} \times \mathbb{R}$, there is a product (say j), with respect to which the marginalization of the demand is permitted. Therefore, there is $(x^{(2)}, p, \delta) \in \mathcal{H}_j$ whose coordinate projection is $(\mathbf{x}^{(2)}, \mathbf{p}, \mathbf{d})$. Hence, one may construct

$$\tilde{\Phi}_j(\mathbf{x}^{(2)}, \mathbf{p}, \mathbf{d}) = \sum_{\mathcal{J} \subseteq \{1, \cdots, J\} \setminus \{j\}} \phi_j \circ \mathcal{M}_{\mathcal{J}}(x^{(2)}, p, \delta) = \int 1\{x_j^{(2)\prime} b^{(2)} + a p_j < -\delta_j\} f_\theta(b^{(2)}, a) d\theta, \tag{B.4}$$

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

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

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 (B.4).

Now define Φ as in (3.20). Then, it follows that

$$\begin{aligned} \Phi(w, u) &= -\int 1\{w'\theta < -u\} f_\theta(b^{(2)}, a) d\theta \\ &= -\int_{-\infty}^{-u} \int_{P_{w,r}} f_\theta(b^{(2)}, a) d\mu_{w,r}(b^{(2)}, a) dr = -\int_{-\infty}^{-u} \mathcal{R}[f_\theta](w, r) dr \ . \end{aligned} \tag{B.5}$$

Taking a derivative with respect to u then yields

$$\frac{\partial \Phi(w, u)}{\partial u} = \mathcal{R}[f_\theta](w, u). \tag{B.6}$$

Note that by Assumption 3.4 $\partial \Phi(w, u)/\partial u$ is either well-defined for all $(w, u) \in \mathbb{H}_+ \times \mathbb{R}$ or only for w in some open neighborhood. In the first case the theorem follows from the injectivity of the Radon transform. In the second case it follow from Lemma B.1. $\square$

The following lemma is used in the proof of Theorem 4.2.

Lemma B.2. *Suppose the Assumptions 2.1 and Condition 4.1 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 $(x^{(2)}, p) = \left(x_1^{(2)}, x_2^{(2)}, p_1, p_2\right) \in$*

$\mathbb{R}^{2k}$ *with* $(x_1^{(2)}, p_1) \neq (x_2^{(2)}, p_2)$ *the function* $\phi : \mathbb{R}^{2k} \times \mathbb{R}^2 \to [0,1]^2$ *defined as*

$$\phi(x_1^{(2)}, x_2^{(2)}, p_1, p_2, d_1, d_2) \equiv \left[\phi_{(0,0)}\left(x_1^{(2)}, x_2^{(2)}, p_1, p_2, d_1, d_2\right), \phi_{(0,1)}\left(x_1^{(2)}, x_2^{(2)}, p_1, p_2, d_1, d_2\right)\right]$$

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 (e.g.* $\tilde{\mathbb{L}} = \{(1,0),(1,1)\}$*).*

Proof of Lemma B.2. 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. Furthermore, the full support of ϵ_1 and ϵ_2 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 $d \in \operatorname{supp}(D)$

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

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 4.2. (a) First, let $\tilde{\mathbb{L}} = \{(1,0),(1,1)\}$. By Condition 4.1 and Lemma B.2, Assumption 2.2 is satisfied. By Assumptions 2.1-2.3 and Theorem 2.1, ψ is identified. Further, the aggregate demand $\{\phi_l, l = (1,0),(1,1)\}$ is identified by Lemma B.2. 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.

Recall that the demand for bundle (0,0) satisfies (4.1). Together with Assumption 3.2 and Lebesgue's theorem the limits

$$\begin{aligned}
\tilde{\Phi}_{(0,0),1}(x_1^{(2)}, p_1, \delta_1) &= -\lim_{\delta_2 \to -\infty} \phi_{(0,0)}(x^{(2)}, p, \delta) = -\int 1\{x_1^{(2)\prime} b^{(2)} + a p_1 + e_1 < -\delta_1\} f_\theta(b^{(2)}, a, e, \Delta) d\theta \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + a p_1 + e_1 < -\delta_1\} f_{\vartheta_1}(b^{(2)}, a, e_1) d\vartheta_1 \\
\tilde{\Phi}_{(0,0),2}(x_2^{(2)}, p_2, \delta_2) &= -\lim_{\delta_1 \to -\infty} \phi_{(0,0)}(x^{(2)}, p, \delta) = -\int 1\{x_2^{(2)\prime} b^{(2)} + a p_2 + e_2 < -\delta_2\} f_\theta(b^{(2)}, a, e, \Delta) d\theta \\
&= -\int 1\{x_2^{(2)\prime} b^{(2)} + a p_2 + e_2 < -\delta_2\} f_{\vartheta_2}(b^{(2)}, a, e_2) d\vartheta_2
\end{aligned}$$

exist and are unique. Note that in both equations Δ and e_1 or e_2 are integrated out. Hence, the first equation connects f_{ϑ_1} to $\tilde{\Phi}_{(0,0),1}$ and the second equation connects f_{ϑ_2} to $\tilde{\Phi}_{(0,0),2}$. Following the argumentation in the proof of Theorem 3.1 yields that f_{ϑ_1} and f_{ϑ_2} are identified.

As a second step we repeat the argument for $\phi_{(1,1)}$. The demand for bundle (1,1) can be written as (4.3). By taking the limits

$$\begin{aligned}
\tilde{\Phi}_{(1,1),1}(x_1^{(2)}, p_1, \delta_1) &= -\lim_{\delta_2 \to -\infty} \phi_{(1,1)}(x^{(2)}, p, \delta) \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 + \Delta < -\delta_1\} f_\theta(b^{(2)}, a, e, \Delta) d\theta \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 + \Delta < -\delta_1\} f_{\eta_1}(b^{(2)}, a, e_1 + \Delta) d\eta_1 \\
\tilde{\Phi}_{(1,1),2}(x_2^{(2)}, p_2, \delta_2) &= -\lim_{\delta_1 \to -\infty} \phi_{(1,1)}(x^{(2)}, p, \delta) \\
&= -\int 1\{x_2^{(2)\prime} b^{(2)} + ap_2 + e_2 + \Delta < -\delta_2\} f_\theta(b^{(2)}, a, e, \Delta) d\theta \\
&= -\int 1\{x_2^{(2)\prime} b^{(2)} + ap_2 + e_2 + \Delta < -\delta_2\} f_{\vartheta_2}(b^{(2)}, a, e_2 + \Delta) d\eta_2
\end{aligned}$$

and following the argument in the proof of Theorem 3.1 the identification of f_{η_1} and f_{η_1} is proven.

(b) With f_{η_j} for $j = 1,2$ the characteristic function $\Psi_{\Delta+\epsilon_j|(\beta^{(2)},\alpha)}$ of $(\Delta_{it} + \epsilon_{ijt})$ conditional on $(\beta_{it}^{(2)}, \alpha_{it})$ is identified as well. With the conditional independence assumption $\Delta_{it} \perp \epsilon_{ijt}|(\beta_{it}^{(2)}, \alpha_{it})$ and $\Psi_{\epsilon_j|(\beta^{(2)},\alpha)}(t) \neq 0$ for almost all $t \in \mathbb{R}$ the densities f_{η_j} and f_{ϑ_j} can be disentangled by the deconvolution:

$$f_{\Delta|\beta^{(2)},\alpha} = \mathcal{F}^{-1}\left(\frac{\Psi_{\Delta+\epsilon_j|(\beta^{(2)},\alpha)}}{\Psi_{\epsilon_j|(\beta^{(2)},\alpha)}}\right),$$

where $\mathcal{F}$ denotes the Fourier transform with respect to Δ. This obviously identifies $f_{\beta^{(2)},\alpha,\Delta}$ as well. If in addition ϵ_{i1t} and ϵ_{i2t} are independent conditional on $(\beta_{it}^{(2)}, \alpha_{it})$, the density of f_θ is identified by

$$f_\theta(b^{(2)}, a, e, \Delta) = f_{\epsilon_1|\beta^{(2)},\alpha}(e_1|b^{(2)}, a)\, f_{\epsilon_2|\beta^{(2)},\alpha}(e_2|\beta^{(2)}, \alpha)\, f_{\beta^{(2)},\alpha,\Delta}(b^{(2)}, a, \Delta)$$

This completes the proof of the thereom. □

Proof of Theorem 4.3. First, let $\tilde{\mathbb{L}} = \{(2,0),(2,1)\}$. By Condition 4.1 and Lemma B.2, Assumption 2.2 is satisfied. By Assumptions 2.1-2.3 and Theorem 2.1, ψ is identified. This implies that the aggregate demand $\{\phi_l, l = (2,0),(2,1)\}$ is identified. 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 4.1, we can take the limits

$$
\begin{aligned}
\tilde{\Phi}_{(0,0)}(x_1^{(2)}, p_1, \delta_1) &= -\lim_{\delta_2 \to -\infty} \phi_{(0,0)}(x^{(2)}, p, \delta) \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 < -\delta_1\} f_\theta(b^{(2)}, a, e, \Delta) d\theta \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 < -\delta_1\} f_{(\beta^{(2)}, \alpha, \epsilon_1)}(b^{(2)}, a, e_1) d\theta \\
\tilde{\Phi}_{(0,1)}(x_1^{(2)}, p_1, \delta_1) &= -\lim_{\delta_2 \to \infty} \phi_{(0,1)}(x^{(2)}, p, \delta) \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 + \Delta_{(1,1)} > -\delta_1\} f_\theta(b^{(2)}, a, e, \Delta_{(1,1)}) d\theta \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 + \Delta_{(1,1)} > -\delta_1\} f_{(\beta^{(2)}, \alpha, \epsilon_1 + \Delta_{(1,1)})}(b^{(2)}, a, e + \Delta_{(1,1)}) d\theta \\
\tilde{\Phi}_{(2,0)}(x_1^{(2)}, p_1, \delta_1) &= -\lim_{\delta_2 \to -\infty} \phi_{(2,0)}(x^{(2)}, p, \delta) \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 + \Delta_{(2,0)} > -\delta_1\} f_\theta(b^{(2)}, a, e, \Delta_{(2,0)}) d\theta \\
&= -\int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 + \Delta_{(2,0)} > -\delta_1\} f_{(\beta^{(2)}, \alpha, \epsilon_1 + \Delta_{(2,0)})}(b^{(2)}, a, e_1 + \Delta_{(2,0)}) d\theta \\
\tilde{\Phi}_{(2,1)}(x_1^{(2)}, p_1, \delta_1) &= -\lim_{\delta_2 \to \infty} \phi_{(2,1)}(x^{(2)}, p, \delta) \\
&= \int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 + \Delta_{(2,1)} - \Delta_{(1,1)} > -\delta_1\} f_\theta(b^{(2)}, a, e, \Delta_{(1,1)}) d\theta
\end{aligned}
$$

$$
= \int 1\{x_1^{(2)\prime} b^{(2)} + ap_1 + e_1 + \Delta_{(2,1)} - \Delta_{(1,1)} > -\delta_1\} f_{(\beta^{(2)}, \alpha, \epsilon_1 + \Delta_{(2,1)} - \Delta_{(1,1)})}(b^{(2)}, a, e_1 + \Delta_{(2,1)} - \Delta_{(1,1)}) d\theta \; .
$$

By the argument in the proof of Theorem 3.1 and the assumption that $(X_{1t}^{(2)}, P_{1t}, D_{1t})$ has a full support, this identifies the joint densities of $(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t})$, $(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t} + \Delta_{i,(1,1),t})$, $(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t} + \Delta_{i,(2,0),t})$, and $(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t} + \Delta_{i,(2,1),t} - \Delta_{i,(1,1),t})$ respectively.

In what follows, the arguments are made conditional on $(\beta_{it}^{(2)}, \alpha_{it})$ unless otherwise noted. By Assumption 4.1 (i), we may disentangle the distribution of ϵ_{i1t} with that of $\Delta_{i,(1,1),t}$, $\Delta_{i,(2,0),t}$, and $\Delta_{i,(2,1),t} - \Delta_{i,(1,1),t}$ respectively by deconvolution as done in the proof of Thoerem 4.2. Thus, the marginal densities of $\Delta_{i,(2,1),t} - \Delta_{i,(1,1),t}$ and $\Delta_{i,(1,1),t}$ are identified. Further, we note that $\Delta_{i,(2,1),t} - \Delta_{i,(1,1),t}$ is a convolution of $\Delta_{i,(2,1),t}$ and $-\Delta_{i,(1,1),t}$. By Assumption 4.1 (ii), Proposition 8 of Carrasco and Florens (2010) applies. Hence, the marginal density of $\Delta_{i,(2,1),t}$ is identified. By Assumption 4.1 (i), $\Delta_{i,(1,1),t} \perp \Delta_{i,(2,0),t} \perp \Delta_{i,(2,1),t}$ conditional on $(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t})$, and each of the marginal densities was identified in the previous step. Therefore, the joint density $f_{(\Delta_{(1,1)}, \Delta_{(2,0)}, \Delta_{(2,1)})|(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t})}$ is identified as the product of the marginal densities. Since the density of $(\beta_{it}^{(2)}, \alpha_{it}, \epsilon_{i1t})$ is identified as well, we may identify the joint density f_{ϑ_1} as $f_{\vartheta_1} = f_{(\Delta_{(1,1)}, \Delta_{(2,0)}, \Delta_{(2,1)})|(\beta^{(2)}, \alpha, \epsilon_1)} f_{(\beta^{(2)}, \alpha, \epsilon_1)}$. f_{ϑ_2} is identified as f_{ϑ_1} by Assumption 4.1 (ii). By Assumption 4.1 (ii) and arguing as in (B.3), f_θ is identified. Given f_θ, all components of ϕ is identified. This completes the proof of the theorem. □

C 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 Dunker et. al. (2014). To give sufficient 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 C.1. *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 C.1 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, Dunker et al. (2014) and Dunker (2015), where D_j now plays the role of the dependent variable.