Identifying the Distribution of Random Coefficients in BLP Demand Models Using One Single Variation in Product Characteristics

Ao Wang

September 2020   **No: 1304**

Warwick Economics Research Papers

Identifying the Distribution of Random Coefficients in BLP Demand Models Using One Single Variation in Product Characteristics*

Ao Wang[†]

September 7, 2020

Abstract

Recent advances on the identification of the Berry, Levinsohn and Pakes (BLP, 1995) random coefficient demand models focus on the structural demand functions. Yet, this does not automatically imply the identification of the distribution of the random coefficients. The latter is often necessary for counterfactuals where the new values of product characteristics do not belong to the support in the factual scenario (e.g. new prices after mergers) or the structural demand functions change (e.g. new products are added). This paper provides novel arguments to identify the distribution of the random coefficients using one single variation in product characteristics. In a leading case where the random coefficients only include a random coefficient on price and individual- and product-specific random intercepts, observing market outcomes at two different price vectors already suffices to identify the distribution of the random coefficients. In theory, these arguments greatly weaken the usual requirements on the regressors or the moments of the random coefficients. In practice, these results are particularly useful when there is little (or limited) variation in product characteristics across markets.

JEL Codes: C4.

Keywords: Identification, Random Coefficients, BLP Model, Demand.

1 Introduction

Since the seminal work of Berry (1994) and Berry et al. (1995) (hereafter BLP), BLP-type models are widely used in the empirical literature of demand. These models typically feature a utility structure with random coefficients that represent individuals' unobserved heterogeneity in price sensitivities and preferences of product characteristics. Oftentimes, the primary goal in identification is to recover the demand functions using market-level data. This enables to identify important objects such as price elasticities and marginal costs at the observed prices. However, only identifying the demand functions may not be sufficient to simulate counterfactuals where the new values of product characteristics are out of the support in the factual scenario, or the structural demand functions change.[1] To simulate these counterfactuals, one has to further identify the distribution of the random coefficients from the market-level data.

This paper provides novel arguments that identify the distribution of the random coefficients in a mixed-logit BLP model of demand. Assuming the identification of the demand functions, the proposed strategy only requires one *single* variation in product characteristics across markets. This requirement is remarkably weaker the often used ones in the existing literature (e.g. special regressors, random coefficients with restricted moments). In a leading case where the random coefficients only include a random coefficient on price and individual- and product-specific random intercepts, observing market outcomes at two different price vectors already suffices to identify the distribution of the random coefficients. This property of robust identification provides an additional argument for using BLP-type models of demand in empirical research.

The identification strategy proceeds in two steps. In the first step, I recover the distribution of the unobserved components in the indirect utilities from the identified demand functions. Leveraging the linear indirect utility structure, this step is to deconvolute the demand functions that are convolutions of multinomial logit and the density function of the unobserved components. In the second step, I aim to recover the joint distribution of the random coefficients in the unobserved components. Because of the linear indirectly utility, this step is to identify the joint distribution of random slopes, which interact with the observed product characteristics, and the random intercepts that

---

[1]One example is merger analysis with the after-merger prices not belonging to the support in the factual case. Another example is product variety analysis where a new (or current) product is added to (or removed from) the choice set.

represent individuals' unobserved perceptions of product qualities. To do so, I assume that the random slopes and the random intercepts are independently distributed. Then, from the first step, I identify the product of the characteristic function of the random slope components (i.e. the interaction between random slopes and the observed characteristics) and that of the random intercepts, conditional on the observed characteristics. By exploiting one single variation in the observed product characteristics, I can difference out the characteristic function of the random intercepts. The identification of the characteristic function of the random slopes follows if the variation shifts product characteristics towards the origin in all directions. Finally, combining this with the identified distribution of the unobserved components, I identify the distribution of the random intercepts.

Related Literature   Recent progress on the identification of demand using aggregate data primarily focuses on the structural demand functions. These progresses include identification arguments using completeness conditions (Berry and Haile, 2014), in a simultaneous system of demand and supply (Matzkin (2008), Berry and Haile (2014, 2018)), in a triangular system (Chesher (2003), Imbens and Newey (2009), D'Haultfœuille and Février (2015), Torgovitsky (2015)), in perturbed utility models of demand (Allen and Rehbeck, 2019), and in models of demand for bundles (Fox and Lazzati (2017), Wang (2019)). However, as pointed out by Fox et al. (2012), in random coefficient models of demand, even if the demand functions are identified, it is still necessary to recover the full distribution of the random coefficients to simulate counterfactuals where new values of product characteristics may not belong to the support in the factual scenario.[2] This paper complements the existing approaches and develops novel arguments to further identify the distribution of the random coefficients.

There is an extensive literature on the identification of random coefficient models. A widely used condition is the existence of a special regressor with large support.[3] Some recent papers relax this requirement and propose strategies that use limited support condition together with restrictions on the location of the support or/and on the moments of

---

[2]See page 206 of Fox et al. (2012).

[3]See Lewbel (2000), Berry and Haile (2009), Fox and Gandhi (2016), Fox and Lazzati (2017), Lewbel and Pendakur (2017), Dunker et al. (2017), Masten (2018) among others.

the random coefficients.[4] The strategy in this paper differs from the existing ones in two aspects. First, it exploits one *single* variation in regressors that interact with the random coefficients to achieve the identification.[5] This strategy significantly alleviates the restrictions on the support of the regressors and the moments of the random coefficients. Second, this strategy requires to recover in advance the structural demand functions at least in an open set. As argued previously, this prerequisite step can be achieved via several well-developed methods in the literature.

The closest papers in the literature are Dunker et al. (2017) and Allen and Rehbeck (2020) whose identification arguments of the distribution of the random coefficients also posit on that the structural demand functions are identified. Dunker et al. (2017) leverages the linear utility structure and employs Radon transformation to identify the distribution of the random coefficients.[6] This approach requires regressors to be continuous and with large support, or limited support but with additional restrictions on the moments of the random coefficients.[7] In contrast, the strategy in this paper applies to both continuous and discrete regressors and only requires a single variation in them. In a perturbed utility model, Allen and Rehbeck (2020) identifies the moments of the random slopes from the (higher-order) derivatives of the demand functions at the origin.[8] The approach of this paper and theirs both employ the condition of independence between random slopes and random intercepts.[9] However, there are two key differences. First, their primary goal is to identify the moments of the random slopes. To identify the distribution, they require further conditions that guarantee that the distribution of the random slopes is uniquely determined by their moments.[10] Differently, the strategy in this paper directly identifies the distribution of the random slopes and therefore does

---

[4]See, for example, Lewbel (2010), Fox et al. (2012), Masten (2018), Chernozhukov et al. (2019), Gaillac and Gautier (2019) and Allen and Rehbeck (2020).

[5]In the setting of control function approach, D'Haultfœuille et al. (2020) proposes a cross condition of the instrument, rather than relying on the exclusion restriction, to identify the distribution of the random coefficients in a linear random coefficient model (see their section 4.1). Similar to the condition in this paper, this cross condition only requires a single variation in the instrument.

[6]See also Hoderlein et al. (2010) and Gautier and Hoderlein (2013) for applications of Radon transformation in identification and estimation of random coefficient models.

[7]See their Assumptions 3.1-3.3.

[8]This strategy is also used by Fox et al. (2012) in the constructive identification arguments.

[9]See their Assumption 1. This condition is fairly standard in both empirical and theoretical literatures. See Chernozhukov et al. (2019) and D'Haultfœuille et al. (2020) among others. Note that oftentimes in empirical research, only random slopes (or random intercepts) are present, while random intercepts (or random slopes) are degenerated, i.e. constants (see Nevo (2000, 2001), Gentzkow (2007), Fox et al. (2012).). This case is also covered by the independence condition.

[10]See Assumption 7 of Fox et al. (2012) for an example of such conditions.

not require such conditions. Second, their strategy to identify the moments requires the support to be around the origin, while the strategy of the current paper does not posit on the special location of the support.

Organisation. Section 2 introduces the model and necessary notations. Section 3 explains the main results of the paper. Section 4 concludes. All examples and proofs can be found in Appendices A-D.

2   Model

Denote market by t.[11] Let $\mathbf{J}_t$ be the set of J_t market-specific products in market t. Without loss of generality, suppose that $\mathbf{J}_t = \mathbf{J}$, i.e. there is no variation in the choice set across markets. Denote the outside option by 0. Individuals in market t can either choose a product $j \in \mathbf{J}$ or the outside option 0. Let $x_{tj} \in \mathbb{R}^K$ denote the vector of observed characteristics of product j in market t. Since the main results of the paper do not necessitate the notational distinction between p_{tj} and x_{tj}, I use x_{tj} to refer to the vector of all observed characteristics of j that also include the price. As in typical BLP models of demand (see Berry and Haile (2014)), I assume the linear index structure in the indirect utilities of products. For individual i in market t, the indirect utility from choosing product j is:

$$
\begin{aligned}
U_{itj} &= x_{tj}\beta_i + \eta_{ij} + \xi_{tj} + \varepsilon_{itj} \\
&= x_{tj}^{(1)}\beta^{(1)} + x_{tj}^{(2)}\beta_i^{(2)} + \eta_{ij} + \xi_{tj} + \varepsilon_{itj} \\
&= [x_{tj}^{(1)}\beta^{(1)} + \eta_j + \xi_{tj}] + [x_{tj}^{(2)}\beta_i^{(2)} + \Delta\eta_{ij}] + \varepsilon_{itj} \\
&= \delta_{tj} + \mu_{itj} + \varepsilon_{itj},
\end{aligned}
$$

where $\delta_{tj} = x_{tj}^{(1)}\beta^{(1)} + \eta_j + \xi_{tj}$ is market t- and product j-specific mean utility, $\mu_{itj} = x_{tj}^{(2)}\beta_i^{(2)} + \Delta\eta_{ij}$ is an individual i-specific utility deviation from δ_{tj}, and ε_{itj} is an idiosyncratic error term. The vector $x_{tj}^{(1)}$ consists of product characteristics that enters U_{itj} with deterministic coefficient(s) $\beta^{(1)}$, i.e. individuals have homogeneous taste on $x_{tj}^{(1)}$, while the vector $x_{tj}^{(2)}$ enters U_{itj} with potentially individual-specific coefficients $\beta_i^{(2)}$, i.e. indi-

---

[11]The definition of market depends on the empirical application. In the case of cross sectional data, one could define markets as different geographic areas; in the case of panel data, they can be defined as different periods. In many cases, market can be defined as a combination of both.

viduals have heterogeneous tastes on $x_{tj}^{(2)}$. The term η_{ij} captures individual i's perception of the quality of product j, with η_j capturing average quality of product j and $\Delta\eta_{ij}$ individual deviation from η_j. Any market-invariant characteristics of product j is then encapsulated in η_j. The term ξ_{tj} is market-product specific demand shock of product j, observed to both firms and individuals but not observed to the econometrician.

For individual i in market t, the indirect utility from choosing the outside option 0 is normalized to

$$U_{it0} = \varepsilon_{it0}.$$

Denote by $\theta_i = (\beta_i^{(2)}, \Delta\eta_i)$ the random coefficients for individual i, where $\Delta\eta_i = (\Delta\eta_{i1}, ..., \Delta\eta_{iJ})$. Suppose that θ_i is distributed according to F. Then, the individual i-specific utility deviation μ_{itj} can be written as $\mu_{itj} = \mu_j(x_{tj}^{(2)}; \theta_i)$. Finally, assume that ε_{it0} and ε_{itj}, $j \in \mathbf{J}$, are i.i.d. Gumbel. We obtain the market share function of product j in market t:

$$s_j(\delta_t; X_t^{(2)}, F) = \int \frac{\exp\{\delta_{tj} + \mu_j(x_{tj}^{(2)}; \theta_i)\}}{1 + \sum_{j' \in \mathbf{J}} \exp\{\delta_{tj'} + \mu_{j'}(x_{tj'}^{(2)}; \theta_i)\}} dF(\theta_i), \tag{1}$$

where $\delta_t = (\delta_{tj})_{j \in \mathbf{J}}$, $X_t^{(2)} = (x_{tj}^{(2)})_{j \in \mathbf{J}} \in \mathbb{R}^{K_2 \times J}$.

In the literature of BLP models of demand, the primary goal of identification is often $s_j(\delta_t; X_t^{(2)}, F)$, for $j \in \mathbf{J}$, as a function of $(\delta_t, X_t^{(2)})$, rather than the distribution F. Obviously, if F is identified, then the market share functions are identified. However, the reverse may not true in general. In the rest of this paper, assuming the identification of the market share functions, I provide additional conditions under which F is identified.

3 Identification of the Distribution F

To ease the exposition, I drop the notation t. Denote the support of $X^{(2)} = (x_j^{(2)})_{j \in \mathbf{J}}$ by $\mathscr{X} \subset \mathbb{R}^{K_2 \times J}$, where K_2 is the dimension of $\beta_i^{(2)}$. To start the discussion, I assume that the market share functions are identified at least in an open set.

Assumption 1. *For any $X^{(2)} \in \mathscr{X}$ and any $j \in \mathbf{J}$, $s_j(\delta; X^{(2)}, F)$ is identified in $\mathscr{D} \ni \delta$, where $\mathscr{D}$ is an open set in $\mathbb{R}^J$.*

Remark 1. *Assumption 1 can be implied by a price-setting game with J supply-side variables. Take cost shifters $c = (c_j)_{j \in \mathbf{J}}$ for example. Because the price vector p (a row*

vector of $X^{(2)}$) is the outcome of the price-setting game, it is typically a function of (δ, c). Then, for a given p, the pricing equation defines a relationship between c and δ. Under suitable regularity conditions, this relationship allows to generate variations of δ in an open subset of $\mathscr{D}$, as required in Assumption 1.

The identification of the market share functions in Assumption 1 can be achieved via several well-developed approaches. In general, without further assumptions, it does not imply the identification of F. However, it is already sufficient for the identification of the distribution of $\mu_i = (\mu_{ij})_{j \in \mathbf{J}}$ conditional on $X^{(2)} \in \mathscr{X}$, as stated in the following theorem:

Theorem 1. *Suppose that Assumption 1 holds. Then, for any $X^{(2)} \in \mathscr{X}$, the distribution of $\mu_i | X^{(2)}$ is identified.*

Proof. See Appendix A. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Remark 2. *If $K_2 = 0$, i.e. $\beta_i^{(2)}$ is degenerated and therefore $\mu_i = \Delta\eta_i$, then Theorem 1 already implies the identification of F.*

Remark 3. *In a more general demand model, Berry and Haile (2014) proves the identification of the distribution of μ_i.[12] Their arguments rely on a condition that the support of μ_i is included in that of the price vector. Differently, Theorem 1 only requires that δ vary in an open set and does not impose any restriction on the support of μ_i.*

We now continue to identify F. We start with a leading case $K_2 = 1$, i.e. $\beta_i^{(2)}$ is a scaler.

3.1   Leading case: $K_2 = 1$

The next Assumption provides a set of sufficient conditions:

Assumption 2.

- *(Single Variation) There exist $X^{(2)}, Y^{(2)} \in \mathscr{X}$, such that for some $j \in \mathbf{J}$, $|x_j^{(2)}| \neq |y_j^{(2)}|$.*

- *(Independence) $\beta_i^{(2)}$ and $\Delta\eta_i$ are independent.*

---

[12]See their section 4.2 on page 1764.

Remarkably weaker than usual support conditions in the literature, the variation condition in Assumption 2 only requires one *single* variation in $X^{(2)}$. In particular, it is implied by any continuous $\mathscr{X}$, or non-singleton discrete $\mathscr{X} \subset \mathbb{R}^J_+$ (e.g. two different price vectors). This condition is motivated by the practical issue that product characteristics may not always change much (or in a limited way) across markets. For this single variation, we require that for some $j \in \mathbf{J}$, $|x^{(2)}_j| \neq |y^{(2)}_j|$, i.e. for some $j \in \mathbf{J}$ the variation shifts characteristics towards the origin.

The independence condition in Assumption 2 is often used in the theoretical literature and also a popular specification in empirical research. In general, it is not necessary when the variation in $X^{(2)}$ is rich enough.[13] However, as shown in Appendix B, without this condition, F may not be identified only using the single variation condition in Assumption 2.

Theorem 2. *Suppose that $K_2 = 1$ and Assumptions 1-2 hold. Then, F is identified.*

Proof. See Appendix C. □

3.2 General case: $K_2 > 1$

For more general cases $K_2 > 1$, i.e. $\beta^{(2)}_i$ is multi-dimensional, the identification of F can be achieved under a similar assumption to Assumption 2:

Assumption 3.

- *(Single Variation) There exists $X^{(2)}, Y^{(2)} \in \mathscr{X}$ and $M \in \mathbb{R}^{J \times J}$, such that $X^{(2)}$ and $Y^{(2)}$ are of full-column rank, $Y^{(2)} = MX^{(2)}$, and the absolute values of the eigenvalues of M are strictly smaller than 1.*

- *(Independence) $\beta^{(2)}_i$ and $\Delta\eta_i$ are independent.*

The independence condition is the same as that in Assumption 2. The single variation condition in Assumption 3 is also along the lines of in Assumption 2 and only requires one variation in the product characteristics matrix. However, the requirement in the case of $K_2 > 1$ is stronger than that in the case of $K_2 = 1$: $X^{(2)}$ is "closer" to the origin than $Y^{(2)}$ in all directions defined by the eigenvectors of M. As in the leading case, the single

---

[13]See the identification analysis of Ichimura and Thompson (1998) and Gautier and Kitamura (2013) for arguments that do not require this independence condition.

variation condition in Assumption 3 is weaker than the support conditions often used in the literature. For example, if the local support condition, i.e $\mathscr{X}$ is an open neighborhood of $X^{(2)}$, holds, then there exists $0 < \lambda < 1$, such that we can define $Y^{(2)} = \lambda \mathbf{I}_{J \times J} X^{(2)}$ and $Y^{(2)} \in \mathscr{X}$. Finally, the full-column rank requirement on $X^{(2)} \in \mathbb{R}^{J \times K_2}$ (or equivalently the full-row rank requirement on $X^{(2)\mathrm{T}}$) guarantees that any vector $v \in \mathbb{R}^{K_2}$ can be expressed by a linear combination of the column vectors of $X^{(2)\mathrm{T}}$.

Theorem 3. *Suppose that $K_2 > 1$, and Assumptions 1 and 3 hold. Then, F is identified.*

See Appendix D for the proof.

4   Conclusion

In this paper, assuming the identification of demand functions, I propose a novel strategy to identify the distribution of the random coefficients in a mixed-logit BLP model of demand. The strategy only requires one single variation in the observed product characteristics that interact with the random coefficients. Compared to the existing literature, this approach does not rely on the existence of a special regressor with large support. This feature is particularly convenient in applications where the value of product characteristics does not vary much (or in a limited way) across markets. Moreover, this strategy does not impose restrictions on the moments of the random coefficients and allow for any distribution, as long as the random slopes and the random intercepts are independently distributed.

Appendix

A   Proof of Theorem 1

Denote the distribution function of $\mu_i = (\mu_j(x_j^{(2)}; \theta_i))_{j \in \mathbf{J}}$ conditional on $X^{(2)}$ by $G_{\mu|X^{(2)}}(\cdot)$. Then, we obtain that for any $j \in \mathbf{J}$,

$$
\begin{aligned}
s_j(\delta; G_{\mu|X^{(2)}}) &= s_j(\delta; X^{(2)}, F) \\
&= \int \frac{\exp\{\delta_j + \mu_{ij}\}}{1 + \sum_{j' \in \mathbf{J}} \exp\{\delta_{j'} + \mu_{ij'}\}} dG_{\mu|X^{(2)}}(\mu_i)
\end{aligned}
$$

is identified for all $\delta \in \mathscr{D}$. Suppose that there exist $G'_{\mu|X^{(2)}}(\cdot)$ such that $s_j(\delta; G_{\mu|X^{(2)}}) = s_j(\delta; G'_{\mu|X^{(2)}})$ for any $\delta \in \mathscr{D}$. In what follows, we prove $G_{\mu|X^{(2)}} = G'_{\mu|X^{(2)}}$ in three steps.

Step 1.

Lemma 1. *Suppose that Assumption 1 holds. Then, for any $X^{(2)} \in \mathscr{X}$ and $j \in \mathbf{J}$, $s_j(\delta; G_{\mu|X^{(2)}}) = s_j(\delta; G'_{\mu|X^{(2)}})$ for $\delta \in \mathbb{R}^J$.*

Remark 4. *When the price coefficient is homogeneous across individuals, the utility structure of model (1) satisfies Assumption 5 in section 4.2 of Berry and Haile (2014). Consequently, keeping other product characteristics fixed, any price change can be equivalently expressed via the change in δ. Then, the change in consumer welfare due to price change is already identified as long as the corresponding path of δ is included in $\mathscr{D}$. Lemma 1 enhances their result in mixed-logit models of demand and already allows to identify consumer welfare change due to any price change (and therefore any path of δ in $\mathbb{R}^J$), without identifying F. This is due to the real-analytic property of demand system (1) with respect to δ_t.*[14]

Proof. According to Theorem 2 (Real Analytic Property) of Iaria and Wang (2019), $s_j(\delta; G_{\mu|X^{(2)}})$ and $s_j(\delta; G'_{\mu|X^{(2)}})$ are both real analytic with respect to δ in $\mathbb{R}^J$. Then, $s_j(\delta; G_{\mu|X^{(2)}}) - s_j(\delta; G'_{\mu|X^{(2)}})$ is also real analytic with respect to δ in $\mathbb{R}^J$. According to Assumption 1, $s_j(\delta; G_{\mu|X^{(2)}}) - s_j(\delta; G'_{\mu|X^{(2)}}) = 0$ in open set $\mathscr{D}$. Then, $s_j(\delta; G_{\mu|X^{(2)}}) - s_j(\delta; G'_{\mu|X^{(2)}}) = 0$ for any $\delta \in \mathbb{R}^J$. $\qquad\square$

Because of Lemma 1, we obtain that for any $\delta \in \mathbb{R}^J$, $\frac{\partial^J s_0(\delta; G_{\mu|X^{(2)}})}{\prod_{j=1}^J \partial \delta_j} = \frac{\partial^J s_0(\delta; G'_{\mu|X^{(2)}})}{\prod_{j=1}^J \partial \delta_j}$.

---

[14]Some papers in the literature have also employed this property in the identification and estimation of mixed-logit models of demand. See Fox et al. (2012), il Kim (2014), Iaria and Wang (2019), Wang (2019).

Equivalently,

$$\frac{\partial^J s_0(\delta; G_{\mu|X^{(2)}})}{\prod_{j=1}^J \partial \delta_j} - \frac{\partial^J s_0(\delta; G'_{\mu|X^{(2)}})}{\prod_{j=1}^J \partial \delta_j}$$

$$= (-1)^J J! \int \frac{1}{1 + \sum_{j' \in \mathbf{J}} \exp\{\delta_{j'} + \mu_{ij'}\}} \prod_{j=1}^J \frac{\exp\{\delta_j + \mu_{ij}\}}{1 + \sum_{j' \in \mathbf{J}} \exp\{\delta_{j'} + \mu_{ij'}\}} d(G_{\mu|X^{(2)}} - G'_{\mu|X^{(2)}})(\mu_i)$$

$$= (-1)^J J! \int \frac{1}{1 + \sum_{j' \in \mathbf{J}} \exp\{\lambda_{j'}\}} \prod_{j=1}^J \frac{\exp\{\lambda_j\}}{1 + \sum_{j' \in \mathbf{J}} \exp\{\lambda_{j'}\}} d(G_{\mu|X^{(2)}} - G'_{\mu|X^{(2)}})(\lambda_i - \delta)$$

$$= (-1)^J J! \int \phi(\lambda_i) d(G_{\mu|X^{(2)}} - G'_{\mu|X^{(2)}})(\lambda_i - \delta)$$

$$= 0,$$

(A.1)

where $\phi(\lambda) = \frac{1}{1 + \sum_{j' \in \mathbf{J}} e^{\lambda_{j'}}} \prod_{j=1}^J \frac{e^{\lambda_j}}{1 + \sum_{j' \in \mathbf{J}} e^{\lambda_{j'}}}$.

Step 2.

Lemma 2. $\phi \in L^1(\mathbb{R}^J)$.

Proof. First, by transforming λ to $\exp\{\lambda\}$, we obtain:

$$\int \phi(\lambda) d\lambda = \int_{\mathbb{R}_+^J} \left(\frac{1}{1 + \sum_{j' \in \mathbf{J}} y_{j'}} \right)^{J+1} dy$$

$$= \sum_{I = \otimes_{j=1}^J I_j, I_j \in \{[0,1),[1,+\infty)\}} \int_I \left(\frac{1}{1 + \sum_{j' \in \mathbf{J}} y_{j'}} \right)^{J+1} dy.$$

Because there are 2^J possible I's, it then suffices to prove that for any I,

$$\int_I \left(\frac{1}{1 + \sum_{j' \in \mathbf{J}} y_{j'}} \right)^{J+1} dy < \infty.$$

Denote the number of j's such that $I_j = [1, +\infty)$ by k. When $k = 0$, $\int_I (1 + \sum_{j' \in \mathbf{J}} y_{j'})^{-(J+1)} dy < 1$. When $k > 0$, without loss of generality, suppose that $I_j = [1, +\infty)$ for $1 \leq j \leq k$, and

11

$I_j = [0, 1)$ for $k + 1 \le j \le J$. Then,

$$\int_I \left(\frac{1}{1 + \sum_{j' \in \mathbf{J}} y_{j'}} \right)^{J+1} dy \le \underbrace{\int_1^{+\infty} \dots \int_1^{+\infty}}_{k} \left(\frac{1}{\sum_{j'=1}^{k} y_{j'}} \right)^{J+1} dy_1 \dots dy_k$$

$$\le \underbrace{\int_1^{+\infty} \dots \int_1^{+\infty}}_{k} \left(\frac{1}{k \prod_{j'=1}^{k} y_{j'}^{\frac{1}{k}}} \right)^{J+1} dy_1 \dots dy_k$$

$$= \frac{1}{k^{J+1}} \left(\int_1^{\infty} y^{-\frac{J+1}{k}} dy \right)^k$$

$$= \frac{1}{k^{J+1-k}} \left(\frac{1}{J + 1 - k} \right)^k.$$

The transition from the first to the second line is obtained by using $\sum_{j'=1}^{k} y_{j'} \ge k (\prod_{j'=1}^{k} y_{j'})^{1/k}$.

$\square$

Because of Lemma 2, $\phi \in L^1(\mathbb{R}^J)$ and hence its Fourier transformation is well defined. Moreover, note that the right-hand side of (A.1) is a convolution of ϕ and $dG_{\mu|X^{(2)}} - dG'_{\mu|X^{(2)}}$.[15] Consequently,

$$\mathscr{F}(\phi)(v)[\psi_{G_{\mu|X^{(2)}}}(v) - \psi_{G'_{\mu|X^{(2)}}}(v)] = 0 \tag{A.2}$$

for any $v \in \mathbb{R}^J$, where $\mathscr{F}(.)$ denotes Fourier transformation and ψ_G is the characteristic function of distribution G.

Step 3.

Lemma 3. *The set $\{v \in \mathbb{R}^J : \mathscr{F}(\phi)(v) = 0\}$ is of zero Lebesgue measure.*

Combining (A.2) and Lemma 3, we obtain that $\psi_{G_{\mu|X^{(2)}}} = \psi_{G'_{\mu|X^{(2)}}}$ almost everywhere. Because characteristic functions are continuous, then we obtain $\psi_{G_{\mu|X^{(2)}}} = \psi_{G'_{\mu|X^{(2)}}}$ every where and hence $G_{\mu|X^{(2)}} = G'_{\mu|X^{(2)}}$. In the remaining part, we prove Lemma 3.

Proof. Note that it suffices to prove that the real (or the imaginary) part of $\mathscr{F}(\phi)$ is real analytic and not constantly zero. As long as this result is proved, according to Mityagin (2015), the zero set of the non-constant real (imaginary) part of $\mathscr{F}(\phi)$ is of zero Lebesgue measure. As a consequence, the zero set of $\mathscr{F}(\phi)$ is also of zero Lebesgue measure.

We first prove the real and imaginary parts of $\mathscr{F}(\phi)$ are real analytic. It suffices to evaluate $\left| \frac{\partial^L \mathscr{F}(\phi)(y)}{\prod_{j=1}^{J} \partial y_j^{l_j}} \right|$, where $\sum_{j=1}^{J} l_j = L$. Note that:

$$\frac{\partial^L \mathscr{F}(\phi)(y)}{\prod_{j=1}^{J} \partial y_j^{l_j}} = \mathscr{F}(\prod_{j=1}^{J} (-\mathrm{i}\lambda_j^{l_j})\phi)(y),$$

---

[15]Here $dG_{\mu|X^{(2)}} - dG'_{\mu|X^{(2)}}$ is defined as a distribution.

where i is the imaginary unit. We now show that for any $y \in \mathbb{R}^J$,

$$\left| \mathscr{F}(\prod_{j=1}^{J}(-\mathrm{i}\lambda_j^{l_j})\phi)(y) \right| \le 2^J J^L \prod_{j=1}^{J} l_j!$$

First,

$$\left| \mathscr{F}(\prod_{j=1}^{J}(-\mathrm{i}\lambda_j^{l_j})\phi)(y) \right| \le \int \prod_{j=1}^{J} |\lambda_j|^{l_j} \phi(\lambda) d\lambda$$

$$= \int_{\mathbb{R}_+^J} \frac{\prod_{j=1}^{J} |\ln y_j|^{l_j}}{(1 + \sum_{j=1}^{J} y_j)^{J+1}} dy$$

$$= \sum_{I = \otimes_{j=1}^{J} I_j, I_j \in \{[0,1),[1,+\infty)\}} \int_I \frac{\prod_{j=1}^{J} |\ln y_j|^{l_j}}{(1 + \sum_{j=1}^{J} y_j)^{J+1}} dy.$$

Similar to the proof of Lemma 2, we evaluate $\int_I \frac{\prod_{j=1}^{J} |\ln y_j|^{l_j}}{(1+\sum_{j=1}^{J} y_j)^{J+1}} dy$ for each I. Denote the number of j's such that $I_j = [1, +\infty)$ by k. When $k = 0$,

$$\int_I \frac{\prod_{j=1}^{J} |\ln y_j|^{l_j}}{(1 + \sum_{j=1}^{J} y_j)^{J+1}} dy \le \prod_{j=1}^{J} \int_0^1 |\ln y_j|^{l_j} dy_j$$

$$= \prod_{j=1}^{J} \int_0^{+\infty} \lambda_j^{l_j} e^{-\lambda_j} d\lambda_j$$

$$= \prod_{j=1}^{J} l_j!$$

When $k > 0$, without loss of generality, suppose that $I_j = [1, +\infty)$ for $1 \le j \le k$, and $I_j = [0, 1)$ for $k + 1 \le j \le J$. Then,

$$\int_I \frac{\prod_{j=1}^{J} |\ln y_j|^{l_j}}{(1 + \sum_{j=1}^{J} y_j)^{J+1}} dy \le \int_1^{+\infty} \frac{\prod_{j=1}^{k} |\ln y_j|^{l_j}}{(1 + \sum_{j=1}^{k} y_j)^{J+1}} dy_1...dy_k \prod_{j=k+1}^{J} \int_0^1 |\ln y_j|^{l_j} dy_j$$

$$\le \frac{1}{k^{J+1}} \int_1^{+\infty} \prod_{j=1}^{k} (\ln y_j)^{l_j} y_j^{-\frac{J+1}{k}} dy_1...dy_k \prod_{j=k+1}^{J} l_j!$$

$$= \frac{1}{k^{J+1}} \prod_{j=k+1}^{J} l_j! \prod_{j=1}^{k} \int_0^{+\infty} \lambda_j^{l_j} e^{-\frac{J+1-k}{k}\lambda_j} d\lambda_j$$

$$= \frac{1}{k^{J+1}} \left(\frac{k}{J+1-k} \right)^L \prod_{j=1}^{J} l_j!$$

$$\le J^L \prod_{j=1}^{J} l_j!.$$

The transition from the first to the second line is obtained by using $\sum_{j=1}^{k} y_j \ge k(\prod_{j=1}^{k} y_j)^{1/k}$.

Then, summing over 2^J integrals, we obtain:

$$\left| \frac{\partial^L \mathscr{F}(\phi)(y)}{\prod_{j=1}^J \partial y_j^{l_j}} \right| \leq 2^J J^L \prod_{j=1}^J l_j!.$$

Denote the real part of $\mathscr{F}(\phi)(y)$ by $\text{Re}[\mathscr{F}(\phi)](y)$. Then, for any $y \in \mathbb{R}^J$,

$$\left| \frac{\partial^L \text{Re}[\mathscr{F}(\phi)](y)}{\prod_{j=1}^J \partial y_j^{l_j}} \right| \leq 2^J J^L \prod_{j=1}^J l_j!.$$

Note that for y such that $|y - y_0| < J^{-2}$, the Taylor expansion of $\text{Re}[\mathscr{F}(\phi)](y)$ around $y = y_0$ can be controlled by

$$\left| \sum_{L=0}^\infty \frac{1}{L!} \left[\sum_{j=1}^J (y_j - y_{0j}) \frac{\partial}{\partial y_j} \right]^L \text{Re}[\mathscr{F}(\phi)](y_0) \right| \leq \sum_{L=0}^\infty \frac{1}{L!} d^L \sum_{\sum l_j = L} \frac{L!}{\prod_{j=1}^J l_j!} \left| \frac{\partial^L \text{Re}[\mathscr{F}(\phi)](y)}{\prod_{j=1}^J \partial y_j^{l_j}} \right|$$

$$\leq 2^J \sum_{L=0}^\infty (dJ^2)^L$$

$$\leq 2^J \sum_{L=0}^\infty \frac{1}{2^L}.$$

The transition from the first to the second line uses $\sum_{\sum_{j=1}^J l_j = L} 1 \leq J^L$. As a result, the Taylor expansion of $\text{Re}[\mathscr{F}(\phi)](y)$ converges for $|y - y_0| < J^{-2}$. Finally, for $|y - y_0| < 0.5 J^{-2}$,

$$\left| \text{Re}[\mathscr{F}(\phi)](y) - \sum_{L=0}^R \frac{1}{L!} \left[\sum_{j=1}^J (y_j - y_{0j}) \frac{\partial}{\partial y_j} \right]^L \text{Re}[\mathscr{F}(\phi)](y_0) \right|$$

$$\leq \left[\frac{1}{2J^2} \right]^{R+1} \sum_{\sum l_j = R+1} \frac{1}{\prod_{j=1}^J l_j!} \sup_{|y - y_0| < \frac{1}{2J^2}} \left| \frac{\partial^L \text{Re}[\mathscr{F}(\phi)](y)}{\prod_{j=1}^J \partial y_j^{l_j}} \right|$$

$$\leq 2^J \left[\frac{1}{2J^2} \right]^{R+1} J^{2(R+1)}$$

$$\to 0.$$

Consequently, $\text{Re}[\mathscr{F}(\phi)]$ is equal to its Taylor expansion and therefore real analytic. Similarly, we can prove that the imaginary part of $\mathscr{F}(\phi)$ is also real analytic. Moreover, because ϕ is not zero functional, then $\mathscr{F}(\phi)$ is not zero functional. As a result, either the real or the imaginary part of $\mathscr{F}(\phi)$ is not constantly zero. The proof is completed. $\square$

14

B   Non-Identification of F without the Independence Condition

In this Appendix, we provide an example where F is not identified when $\mathcal{X}$ only has two points and $\beta_i^{(2)}$ and $\Delta\eta_i$ are not independent.

Suppose that $J = 1$, i.e. there is only one inside product, and $(\beta_i^{(2)}, \Delta\eta_i)$ follows a centered normal distribution with a covariance matrix Ω. We know that Ω has 3 unknowns: the variance of $\beta_i^{(2)}$, the variance of $\Delta\eta_i$, and their correlation $r \neq 0$. Suppose that the support $\mathcal{X}$ only has two points: $\mathcal{X} = \{x, y\}$ and that the distribution of $\mu_i = x^{(2)}\beta_i^{(2)} + \Delta\eta_i$ is identified conditional on $x^{(2)} = x, y$. Then, μ_i conditional on $x^{(2)}$ follows a centered normal distribution with the variance being $(x^{(2)}, 1)\Omega(x^{(2)}, 1)^{\mathrm{T}}$. We can identify $(x, 1)\Omega(x, 1)^{\mathrm{T}}$ and $(y, 1)\Omega(y, 1)^{\mathrm{T}}$. Without further assumptions, we obtain 2 equations with 3 unknowns. Then, Ω cannot be uniquely determined and therefore the distribution of $(\beta_i^{(2)}, \Delta\eta_i)$ is not identified.

C   Proof of Theorem 2

Because of the independence between $\beta_i^{(2)}$ and $\Delta\eta_i$, it suffices to identify the distributions of $\beta_i^{(2)}$ and $\Delta\eta_i$. According to Assumption 2, without loss of generality, suppose that the first elements of $X^{(2)}$ and $Y^{(2)}$ are different and $|x_1^{(2)}| > |y_1^{(2)}|$. Moreover, because of Theorem 1, the distribution of $\mu_i|X^{(2)}$ (and also $\mu_i|Y^{(2)}$) is identified. In particular, the distribution of $\mu_{i1} = x_1^{(2)}\beta_i^{(2)} + \Delta\eta_{i1}$ conditional on $x_1^{(2)}$ ($y_1^{(2)}$) is identified. Because of the independence condition in Assumption 2, we obtain that

$$\psi_{\mu_1|x_1^{(2)}}(v) = \psi_{\beta^{(2)}}(x_1^{(2)}v)\psi_{\Delta\eta_1}(v),$$

for any $v \in \mathbb{R}$, where $\psi_w(v)$ denotes the characteristic function of random variable w evaluated at v. As a consequence,

$$\begin{aligned}
\psi_{\mu_1|x_1^{(2)}}(v) &= \psi_{\beta^{(2)}}(x_1^{(2)}v)\psi_{\Delta\eta_1}(v), \\
\psi_{\mu_1|y_1^{(2)}}(v) &= \psi_{\beta^{(2)}}(y_1^{(2)}v)\psi_{\Delta\eta_1}(v),
\end{aligned} \tag{C.1}$$

and the left-hand sides (C.1) are identified. Then, the ratio $r(v) = \psi_{\beta^{(2)}}(x_1^{(2)}v)/\psi_{\beta^{(2)}}(y_1^{(2)}v)$ is identified for any $v \in \mathbb{R}$. If $y_1^{(2)} = 0$, then $\psi_{\beta^{(2)}}(y_1^{(2)}v) = 1$ and $\psi_{\beta^{(2)}}(x_1^{(2)}v) = r(v)$. Since $x_1^{(2)} \neq 0$, we then identify $\psi_{\beta^{(2)}}(v)$ for any $v \in \mathbb{R}$. Consequently, the distribution of $\beta^{(2)}$ is identified. If $y_1^{(2)} \neq 0$, since $|x_1^{(2)}| > |y_1^{(2)}|$, then $x_1^{(2)} \neq 0$. Note that for any $v \in \mathbb{R}$,

$$\psi_{\beta^{(2)}}(v) = r\left(\frac{v}{x_1^{(2)}}\right)\psi_{\beta^{(2)}}\left(\frac{y_1^{(2)}}{x_1^{(2)}}v\right). \tag{C.2}$$

Then, by using $\psi_{\beta^{(2)}}(0) = 1$ and $|y_1^{(2)}| < |x_1^{(2)}|$, we can iterate (C.2): for any integer L,

$$\psi_{\beta^{(2)}}(v) = \prod_{l=0}^{L} r\left(\left[\frac{y_1^{(2)}}{x_1^{(2)}}\right]^l \frac{v}{x_1^{(2)}}\right) \psi_{\beta^{(2)}}\left(\left[\frac{y_1^{(2)}}{x_1^{(2)}}\right]^L v\right),$$

$$\psi_{\beta^{(2)}}(v) = \prod_{l=0}^{\infty} r\left(\left[\frac{y_1^{(2)}}{x_1^{(2)}}\right]^l \frac{v}{x_1^{(2)}}\right).$$

Then, $\psi_{\beta^{(2)}}(v)$ for any $v \in \mathbb{R}$ and therefore the distribution of $\beta^{(2)}$ is identified. Finally, to identify the distribution of $\Delta\eta_i$, note that $\mu_i = X^{(2)}\beta_i^{(2)} + \Delta\eta_i$. Given the independence of $\beta_i^{(2)}$ and $\Delta\eta_i$, we obtain: for any $\nu \in \mathbb{R}^J$,

$$\psi_{\mu_i|X^{(2)}}(\nu) = \psi_{\beta^{(2)}}(X^{(2)\mathrm{T}}\nu)\psi_{\Delta\eta}(\nu).$$

Because $\psi_{\mu_i|X^{(2)}}(\nu)$ and $\psi_{\beta^{(2)}}(X^{(2)\mathrm{T}}\nu)$ are identified, then $\psi_{\Delta\eta}(\nu)$ is identified. Consequently, the distribution of $\Delta\eta_i$ is identified. The proof is completed.

D Identification of F when $K_2 > 1$

Note that the distribution of $\mu_i = X^{(2)}\beta_i^{(2)} + \Delta\eta_i$ conditional on $X^{(2)}$ is identified. Given the independence of $\beta_i^{(2)}$ and $\Delta\eta_i$, we then deduce that

$$\psi_{\mu_1|X^{(2)}}(\nu) = \psi_{\beta^{(2)}}(X^{(2)\mathrm{T}}\nu)\psi_{\Delta\eta}(\nu),$$

$$\psi_{\mu_1|Y^{(2)}}(\nu) = \psi_{\beta^{(2)}}(Y^{(2)\mathrm{T}}\nu)\psi_{\Delta\eta}(\nu),$$

and therefore the ratio $r(\nu) = \psi_{\beta^{(2)}}(X^{(2)\mathrm{T}}\nu)/\psi_{\beta^{(2)}}(Y^{(2)\mathrm{T}}\nu)$ is identified for any $\nu \in \mathbb{R}^J$. Because $X^{(2)}$ is of full column rank, then $X^{(2)\mathrm{T}}$ is of full row rank K_2 and $K_2 \leq J$. As a consequence, for any $v \in \mathbb{R}^{K_2}$, there exists some ν such that $X^{(2)\mathrm{T}}\nu = v$. Then,

$$\begin{aligned}
\psi_{\beta^{(2)}}(v) &= \psi_{\beta^{(2)}}(X^{(2)\mathrm{T}}\nu) \\
&= r(\nu)\psi_{\beta^{(2)}}(Y^{(2)\mathrm{T}}\nu) \\
&= r(\nu)\psi_{\beta^{(2)}}(X^{(2)\mathrm{T}}M\nu) \\
&= r(\nu)r(M\nu)\psi_{\beta^{(2)}}(Y^{(2)\mathrm{T}}M\nu) \\
&= \prod_{l=0}^{L} r(M^l\nu)\psi_{\beta^{(2)}}(Y^{(2)\mathrm{T}}M^L\nu).
\end{aligned}$$

Because the absolute values of the eigenvalues of M is strictly smaller than 1, then $|M^L\nu| \to 0$ as $L \to \infty$. Consequently, $\psi_{\beta^{(2)}}(v) = \prod_{l=0}^{\infty} r(M^l\nu)$ and therefore identified. This implies the identification of the distribution of $\beta_i^{(2)}$. Finally, the identification of the distribution of $\Delta\eta_i$

follows from that of its characteristic function of $\Delta\eta_i$: $\psi_{\Delta\eta}(\nu) = \psi_{\mu_1|X^{(2)}}(\nu)/\psi_{\beta^{(2)}}(X^{(2)\mathrm{T}}\nu)$.

References

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

Allen, R. and Rehbeck, J. (2020). Identification of random coefficient latent utility models. *arXiv preprint arXiv:2003.00276*.

Armstrong, T. B. (2016). Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica*, 84(5):1961–1980.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

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

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

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

Berry, S. T. and Haile, P. A. (2018). Identification of nonparametric simultaneous equations models with a residual index structure. *Econometrica*, 86(1):289–315.

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

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, 71(5):1405–1441.

D'Haultfœ uille, X., Hoderlein, S., and Sasaki, Y. (2020). Testing and relaxing the exclusion restriction in the control function approach. *Working Paper*, pages 1–45.

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

Dunker, F., Hoderlein, S., and Kaido, H. (2017). Nonparametric identification of random coefficients in endogenous and heterogeneous aggregated demand models. *cemmap working paper*.

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

Fox, J. T., Kim, K., Ryan, S. P., and Bajari, P. (2012). The random coefficients logit model is identified. *Journal of Econometrics*, 166(2):204–212.

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

Gaillac, C. and Gautier, E. (2019). Adaptive estimation in the linear random coefficients model when regressors have limited variation. *arXiv preprint arXiv:1905.06584*.

Gautier, E. and Hoderlein, S. (2013). Estimating the distribution of treatment effects. *arXiv preprint arXiv:1109.0362*.

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

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

Hoderlein, S., Klemelä, J., and Mammen, E. (2010). Analyzing the random coefficient model nonparametrically. *Econometric Theory*, 26(3):804–837.

Iaria, A. and Wang, A. (2019). Identification and estimation of demand for bundles. *Working Paper*, pages 1–92.

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

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

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

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, 97(1):145–177.

Lewbel, A. (2010). Identification of limited dependent variable models using a special regressor with limited support. *unpublished manuscript*.

Lewbel, A. and Pendakur, K. (2017). Unobserved preference heterogeneity in demand using generalized random coefficients. *Journal of Political Economy*, 125(4):1100–1148.

Masten, M. A. (2018). Random coefficients on endogenous variables in simultaneous equations models. *The Review of Economic Studies*, 85(2):1193–1250.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. *Econometrica*, 76(5):945–978.

Mityagin, B. (2015). The zero set of a real analytic function. *arXiv preprint arXiv:1512.07276*.

Nevo, A. (2000). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics*, pages 395–421.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.

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

Wang, A. (2019). A blp demand model of product-level market shares with complementarity. *Working Paper*, pages 1–72.