YOGURTS CHOOSE CONSUMERS? ESTIMATION OF RANDOM-UTILITY MODELS VIA TWO-SIDED MATCHING

ODRAN BONNET[♭], ALFRED GALICHON[†], YU-WEI HSIEH[♯], KEITH O'HARA[♢], AND MATT SHUM[§]

Abstract. The problem of *demand inversion* – a crucial step in the estimation of random utility discrete-choice models – is equivalent to the determination of stable outcomes in two-sided matching models. This equivalence applies to random utility models that are not necessarily additive, smooth, nor even invertible. Based on this equivalence, algorithms for the determination of stable matchings provide effective computational methods for estimating these models. For non-invertible models, the identified set of utility vectors is a lattice, and the matching algorithms recover sharp upper and lower bounds on the utilities. For invertible models, our matching approach facilitates estimation of models that were previously difficult to estimate, such as the pure characteristics model. An empirical application to voting data from the 1999 European Parliament elections illustrates the good performance of our matching-based demand inversion algorithms in practice.

Keywords: random utility models, demand inversion, two-sided matching, discrete-choice demand models, partial identification, pure characteristics model

JEL Classification: C51, C60

Date: July 27, 2020 (First draft: 9/2015). Galichon gratefully acknowledges funding from a grant NSF DMS-1716489, and from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no. 295298. We thank Jeremy Fox, Xavier d'Haultefoeuille, Lars Nesheim, Ariel Pakes, and participants in seminars at Caltech, UC-Davis, Johns Hopkins, Rochester (Simon), Stanford, UNC, USC, Yale, the NYU CRATE conference, the Banff Applied Microeconomics Conference, the SHUFE Econometrics Conference, the Toronto Intersections of Econometrics and Applied Micro Conference, WARP (Workshop of Applications of Revealed Preference) webinar, and UCL-Vanderbilt Conference on Econometrics and Models of Strategic Interactions for helpful comments. Alejandro Robinson-Cortes provided excellent research assistance.

[♭]Sciences Po; odran.bonnet@sciencespo.fr

[†]New York University; ag133@nyu.edu

[♯]University of Southern California; yuwei.hsieh@usc.edu

[♢]Amazon.com; keith.ohara@nyu.edu. O'Hara's contributions to the paper reflect work done prior to his joining Amazon.

[§]California Institute of Technology; mshum@caltech.edu .

1. Introduction

Discrete choice models play a tremendous role in applied work in economics. In these models, an agent i characterized by a utility shock $\varepsilon_i \in \Omega$ must choose from a finite set of alternatives $j \in \mathcal{J}$ in order to maximize her utility. The random utility framework pioneered by McFadden et al. (1978) assumes that the utility $\mathcal{U}_{\varepsilon_i j}(\delta_j)$ that agent i gets from alternative j depends on δ_j, a systematic utility level associated with alternative j, which is identical across all agents, and a realization ε_i of agent i's random utility shocks. The agent chooses the alternative yielding maximal utility:

$$\max_{j \in \mathcal{J}} \left\{ \mathcal{U}_{\varepsilon_i j}(\delta_j) \right\}. \tag{1.1}$$

We focus on parametric random utility models, where the function $\mathcal{U} : (\varepsilon, j, \delta) \in \Omega \times \mathcal{J} \times \mathbb{R} \mapsto \mathcal{U}_{\varepsilon j}(\delta) \in \mathbb{R}$ as well as the distribution of ε in the population, denoted P, are known to the researcher. Particular instances of these models are Additive Random Utility Models (hereafter ARUMs), including Logit or Probit models, where $\Omega = \mathbb{R}^{\mathcal{J}}$, $\mathcal{U}_{\varepsilon_i j}(\delta_j) = \delta_j + \varepsilon_{ij}$, and ε follows a Gumbel or Gaussian distribution. However, our results extend to the more general class of Non-Additive Random Utility Models (NARUMs), in which $\mathcal{U}_{\varepsilon_i j}(\delta_j)$ is not (quasi-)linear in δ_j.

Under an assumption guaranteeing that agents are not indifferent between any pair of alternatives (see Assumption 2 below), we can define the vector-valued *demand map* $\sigma(.)$, the j-th component ($j \in \mathcal{J}_0$) of which is defined as the probability that alternative j dominates all the other ones, given the vector of systematic utilities $(\delta_j)_{j \in \mathcal{J}}$:

$$\sigma_j(\delta) = P\left(\varepsilon : \mathcal{U}_{\varepsilon j}(\delta_j) \geq \mathcal{U}_{\varepsilon j'}\left(\delta_{j'}\right), \forall j' \in \mathcal{J}\right). \tag{1.2}$$

The main focus of the paper pertains to *demand inversion*: given a vector of observed market shares $(s_j)_{j \in \mathcal{J}}$, how can one characterize and compute the full set of utility vectors $(\delta_j)_{j \in \mathcal{J}}$ such that $s = \sigma(\delta)$ – that is, which rationalizes the observed market shares? Additionally, a partial identification situation may arise, as the identified set of vectors δ that solve the demand inversion problem may not necessarily be a single point.

1.1. Contribution.

Contribution. We establish a new equivalence principle between the problem of demand inversion and the problem of stable matchings in two-sided models with Imperfectly Transferable Utility (ITU). More precisely, we show that a discrete choice model can always be interpreted as a two-sided matching market where consumers and alternatives are viewed as firms and workers; and that the *demand inversion problem*, that is the identification of utility vectors $(\delta_j)_{j \in \mathcal{J}}$ can be reformulated as the *equilibrium problem* of determining competitive wages in the corresponding matching market. In other words, the identified set of solution vectors δ coincides with the set of equilibrium wages in the matching market. This equivalence implies two important contributions:

(1) **Characterization of the identified set of δ.** The equivalence to the matching equilibrium implies that the identified set of vectors δ_j is a *lattice.*, from which one can construct a very simple data-dependent test for point-identification.[1] As such, if the greatest element of the lattice coincides with its smallest element, then the utility index δ is point-identified, which implies a simple data-dependent assessment of point-identification.[2] Thus, our approach bypasses the need of verifying *a priori* whether the parameters of a given model are point of partial-identified—a non-trivial exercise in many cases.

(2) **Computation of the identified set of δ.** Our matching approach has two key features. First, the matching equivalence allows the utilization of several high-performance matching algorithms, for which the convergence properties are well-studied. The use of these matching algorithms for estimating random utility models is new; moreover, they can readily handle partial-identification situations in which multiple values of δ rationalize the observed market shares. Second, these matching-based algorithms do not require the computation of the demand (market-share) mapping. This is important an specific models, such as the *pure characteristics model* (Berry and Pakes (2007)), are notorious for their non-smooth market-share mappings. Indeed, using our approach, the demand inversion problem for the pure characteristics model becomes a well-behaved *convex program* (see Section 5 below).

Demand inversion is a crucial intermediate step for estimating aggregate discrete-choice models of product-differentiated markets; see, e.g., Berry (1994) and Berry, Levinsohn, and Pakes (1995) (BLP). It also plays an important role in two-step estimation procedures for dynamic discrete-choice models (including Hotz and Miller (1993), Aguirregabiria and

---

[1]A lattice is a partially ordered set that contains the meet and the join of each pair of its element. For the purposes of this paper, lattices are subsets of vectors in Euclidean space and, for any given pair of vectors, the meet (resp. join) is just the vector containing the componentwise infimum (resp. supremum). For additional discussion of lattices in matching theory, consult (Roth and Sotomayor 1992). Relatedly, Jia-Barwick (2008) exploits the lattice structure of equilibria in supermodular games to estimate a large multi-market entry game between discount retailers.

[2]Khan, Ouyang, and Tamer (2016) call this an "adaptive" property.

Mira (2002), Bajari, Benkard, and Levin (2007), Arcidiacono and Miller (2011), Kristensen, Nesheim, and de Paula (2014)).

To date, the existing literature has not provided a general characterization of the identified set $\sigma^{-1}(s)$, defined as the set of utility vectors (δ_j) which rationalize a vector of market shares (s_j). This paper is the first to consider situations when the identified utility set $\sigma^{-1}(\{s\})$ it is non-singleton. In addition, most of the papers cited above provide little guidance on computing the identified set. As Berry and Haile (2015, p. 10) underline, "(...) the invertibility result of Berry, Gandhi, and Haile (2013) is not a characterization (or computational algorithm) for the inverse". Besides a handful of models,[3] there are no well-established procedures for demand-inversion in general (non-additive) random utility models with arbitrary error distributions. Our paper aims to fill this gap.

This paper also follows upon a set of recent papers which have reformulated the problem of demand inversion in ARUMs as an optimal transport problem, using the tools of convex duality. This approach was pioneered by Galichon and Salanié (2015), and was extended to ARUMs with possibly noncontinuous distributions of unobserved heterogeneity by Chiong, Galichon, and Shum (2016), and to continuous choice problems by Chernozhukov, Galichon, Henry, and Pass (2019). However, these papers do not cover nonadditive random utility models, and do not characterize the structure of the identified set. A more in-depth discussion of the related literature is provided below in Section 8.

1.2. **Organization.** Section 2 introduces the general random utility framework which is the focus of this paper and provides examples. Section 3 presents our main equivalence result between NARUMs and two-sided matching problems, and discusses the lattice structure of the identified utility set. Based on the equivalence result, in Section 4 we introduce several matching-based algorithms which can solve a wide variety of random utility models. Section 5 contains two simulation investigations of the algorithms, including the pure characteristics model. Section 6 utilizes our matching approach to estimate a spatial voting model using electoral data from the 1999 European Parliament elections. Section 7 collects additional theoretical results on existence, and uniqueness, and Section 8 provides a detailed literature review. Section 9 concludes. All proofs are collected in the appendix.

[3]These include the logit, nested logit, and random-coefficient logit models (Berry (1994), Berry, Levinsohn, and Pakes (1995), Dubé, Fox, and Su (2012)).

2. The framework

2.1. **Basic assumptions.** Let $\mathcal{J}_0 = \mathcal{J} \cup \{0\}$ be a finite set of alternatives, where $j = 0$ denotes a special alternative which serves as a benchmark (see section 2.2 below). The agent's program is thus

$$u_{\varepsilon_i} = \max_{j \in \mathcal{J}_0} \{\mathcal{U}_{\varepsilon_i j}(\delta_j)\}, \tag{2.1}$$

where u_{ε_i} is the indirect utility of an agent with shock ε_i. The utility agent i derives from alternative j depends on the systematic utility vector δ_j associated with this alternative, and on the realization ε_i of this agent's utility shock. We will work under two assumptions:

Assumption 1 (Regularity of $\mathcal{U}$). *Assume (Ω, P) is a Borel probability space and for every $\varepsilon \in \Omega$, and for every $j \in \mathcal{J}_0$:*

(a) the map $\varepsilon \mapsto (\mathcal{U}_{\varepsilon j}(\delta_j))_{j \in \mathcal{J}_0}$ is measurable, and

(b) the map $\delta_j \mapsto \mathcal{U}_{\varepsilon j}(\delta_j)$ is increasing from $\mathbb{R}$ to $\mathbb{R}$ and continuous.

Assumption 2 (No indifference). *For every distinct pair of indices j and j' in $\mathcal{J}_0$, and for every pair of scalars δ and δ',*

$$P\left(\varepsilon \in \Omega : \mathcal{U}_{\varepsilon j}(\delta) = \mathcal{U}_{\varepsilon j'}(\delta')\right) = 0.$$

These two assumptions are standard in the literature, and are automatically satisfied in ARUMs. Assumption 1 (a) is a standard measurability condition, and (b) is often invoked in the literature on non-separable models (see, e.g., Matzkin (2007)).

Assumption 2 rules out indifference (precisely, an event of measure zero) between two alternatives, and is maintained in practically all the applied discrete choice literature. In appendix A, we show that the results in this paper hold even without Assumption 2, albeit at the greater notational expense of introducing set-valued functions. Hence, for simplicity, in the main text we maintain Assumption 2 as it suffices for our purposes.

Under Assumption 2, the demand of alternative j (defined in (1.2) above) corresponds to the fraction of consumers who prefer weakly *or* strictly alternative j to any other one:

$$\sigma_j(\delta) := P\left(\varepsilon \in \Omega : \mathcal{U}_{\varepsilon j}(\delta_j) \geq \max_{j' \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j'}(\delta_{j'})\right) = P\left(\varepsilon \in \Omega : \mathcal{U}_{\varepsilon j}(\delta_j) > \max_{j' \in \mathcal{J}_0 \setminus \{j\}} \mathcal{U}_{\varepsilon j'}(\delta_{j'})\right). \tag{2.2}$$

Importantly, the uniqueness of the vector of market shares associated with a given utility vector δ does not imply that the demand inversion problem has a unique solution. There may be multiple vectors δ such that $\sigma(\delta) = s$. Under Assumptions 1 and 2, the vector of market shares $s = \sigma(\delta)$ is a probability vector on $\mathcal{J}_0$, which prompts us to introduce $\mathcal{S}_0$, the set of such probability vectors as

$$\mathcal{S}_0 := \left\{ s \in \mathbb{R}_+^{\mathcal{J}_0} : \sum_{j \in \mathcal{J}_0} s_j = 1 \right\}.$$

We formalize the definition of the demand map.

Definition 1 (Demand map)**.** Under Assumption 1 and 2, the *demand map* is the map $\sigma : \mathbb{R}^{\mathcal{J}_0} \to \mathcal{S}_0$ defined by expression (2.2).

2.2. **Normalization.** Any discrete choice model requires some normalization, because the choice probabilities result from the comparison of the relative utility payoffs from each alternative. Throughout the paper we normalize the systematic utility associated to the default alternative to zero:

$$\delta_0 = 0, \tag{2.3}$$

and we use

$$\tilde{\sigma} : \mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{J}} \tag{2.4}$$

to denote the map induced by this normalization.

In the special ARUM case where $\mathcal{U}_{\varepsilon ij}(\delta_j) = \delta_j + \varepsilon_{ij}$, imposing normalization (2.3) is innocuous because that the vector of systematic utilities (δ_j) yields the same choice problem as the vector $(\delta_j + c)$ where c is a constant; hence, for ARUMs, any normalization will yield the same identified utility vectors δ up to an additive constant. However, this is no longer true in nonadditive models, for which the normalization (2.3) entails some loss of generality. We explore this below in Section 5.

2.3. **Examples.** Next, we consider several examples of random utility models falling within our framework. Since we assume that market shares are generated by the mapping in Eq. (2.2), we implicitly assume that the random element ε is independently and identically distributed across all consumers in the market. However, we make no restrictions on ε across alternatives: as the examples below show, ε can be individual-specific, choice-specific, or some combination of the two.

Example 2.1 (ARUM)**.** In the additive random utility model (ARUM) one sets, $\Omega = \mathbb{R}^{\mathcal{J}_0}$, so that P is a probability distribution on $\mathbb{R}^{\mathcal{J}_0}$, and

$$\mathcal{U}_{\varepsilon ij}(\delta_j) = \delta_j + \varepsilon_{ij}.$$

There are several well-known instances of ARUMs:

Logit model: if P is the distribution of a vector of size $|\mathcal{J}_0|$ of i.i.d. type 1-Extreme value random variables, then the demand map is given by $\sigma_j(\delta) = \exp(\delta_j) / \left(\sum_{j' \in \mathcal{J}_0} \exp(\delta_{j'})\right)$. Berry (1994) shows that this demand map is invertible: δ_j is point identified by the following "log-odds ratio" formula:

$$\delta_j = \log(s_j / s_0). \tag{2.5}$$

Pure characteristics model: In this model, consumers value product j only through its measurable characteristics $x_j \in \mathbb{R}^d$, a vector of dimension d associated to each alternative j, and the utility shock vector ε_i is such that

$$\varepsilon_{ij} = \nu_i^\intercal x_j = \sum_{k=1}^{d} \nu_i^k x_j^k \tag{2.6}$$

where ν_i is consumer i's vector of taste-shifters, drawn from a distribution P_ν on $\mathbb{R}^d$. In this case, there is no closed-form expression for the demand map. Berry and Pakes (2007) (p. 1193) underline that this model is appealing on theoretical grounds, but infrequently used in empirical work, arguably due to computational challenges.[4] Our matching approach yields computationally tractable ways to estimate it, which we highlight in the simulations and empirical application in sections 5 and 6.

Random coefficient logit model: In the random coefficient logit model popularized by BLP and McFadden and Train (2000), the random-utility shock is given by:

$$\varepsilon_{ij} = \nu_i^\intercal x_j + \zeta_{ij}.$$

This is the sum of two independent terms: one logit term ζ_{ij} and one pure characteristics term $\nu_i^\intercal x_j$.

Example 2.2 (Risk aversion)**.** Consider a market where consumers are not fully aware of the attributes of a product at the time of purchase. This may characterize consumers' choices in online markets, where they have no opportunity to physically examine the goods under consideration. Let ε_i denote the relative risk aversion parameter (under CRRA utility), and that the price of good j is p_j. Choosing option j yields a consumer surplus of $\delta_j - p_j + \eta_j$ where $\log \eta_j \sim N(0,1)$ is a quality shock unobservable at the time of the

[4]See Song (2007) and Nosko (2010) for two empirical applications of the pure characteristics demand model. Pang, Su, and Lee (2015) provide computational algorithms for estimating this model.

purchase, and δ_j is the willingness to pay (in dollar terms) associated to alternative j. At the time of the purchase, the consumer's expected utility is

$$\mathcal{U}_{\varepsilon_i j}(\delta_j) = \mathbb{E}_{\eta_j}\left[\frac{(\delta_j - p_j + \eta_j)^{1-\varepsilon_i}}{1-\varepsilon_i}\right],$$

where the expectation is taken over η_j holding ε_i constant. These kind of models are typically non-additive in ε.[5]

Example 2.3 (Vertical differentiation model)**.** In the classic vertical differentiation demand framework,[6] household i obtains utility from brand j equal to

$$\delta_j \theta_i - p_j, \quad \forall j.$$

Here δ_j is interpreted as the quality of brand j, while the nonlinear random utility shock θ_i measures household i's willingness-to-pay for quality. Below, in Section 5, we will consider a numerical example based on this framework which is non-additive and not point identified.

Example 2.4 (Retirement decision)**.** Assume that utility from consumption basket z is $V(z) + \delta_0$ if agent i is not retiring (option $j = 0$), in which case she gets labour income y_0, and $V(z) + \delta_1$ if retiring (option $j = 1$), in which case she gets pension income y_1. Agent i's non-labour income is ε_i. Then

$$\begin{aligned}
\mathcal{U}_{\varepsilon_i 0}(\delta_0) &= \max_{z \in R^d}\left\{V(z) + \delta_0 : z'p \le y_0 + \varepsilon_i\right\} \\
\mathcal{U}_{\varepsilon_i 1}(\delta_1) &= \max_{z \in R^d}\left\{V(z) + \delta_1 : z'p \le y_1 + \varepsilon_i\right\}.
\end{aligned}$$

Example 2.5. (Investments with taxes) While our framework requires each δ_j to be scalar, the nature of the consumer heterogeneity ε can be quite general. To illustrate this, we consider an example with two components of consumer heterogeneity: $\varepsilon = (\varepsilon^1, \varepsilon^2)$. While ε^2 denote the usual additive random-utility errors, $\varepsilon \in \{0, 1\}$ is binary, and denotes a consumer's tax-exampt status, which is assumed unobservable to the researcher. We consider an individual who is considering investing among $j \in \mathcal{J}$ projects. Letting $\delta_j + \varepsilon_j^2$ denote project j's pre-tax earnings, and τ the tax rate, the return from project j is

$$\mathcal{U}_{\varepsilon j}(\delta_j) = \varepsilon^1\left(\delta_j + \varepsilon_j^2\right) + \left(1 - \varepsilon^1\right)(1-\tau)\left(\delta_j + \varepsilon_j^2\right).$$

[5]See Cohen and Einav (2007) and Apesteguia and Ballester (2014) for examples.

[6]See, among others, Prescott and Visscher (1977) and Bresnahan (1981). Berry and Pakes (2007) extends this framework to the multivariate case.

3. Equivalence of Discrete-Choice and Two-Sided Matching

In this section, we show a central result of this paper; namely, an equivalence between discrete-choice models and two-sided matching problems. This equivalence is noteworthy as discrete choice problems are traditionally considered to be "one-sided" problems. However, we will demonstrate that they are equivalent to a two-sided "marriage problem" between consumers and yogurts, where both sides of the market must assent to be matched. It immediately follows from this equivalence that the demand inversion problem can be equivalently formulated as solving for equilibrium payoffs from the corresponding two-sided matching problem.

We begin by formally defining the object of interest for demand inversion, which is to recover the identified utility set:

Definition 2 (Identified utility set)**.** Given a demand map $\tilde{\sigma}$ defined as in (2.4) where Assumptions 1 and 2 are met, and given a vector of market shares s that satisfies $s_j > 0$ and $\sum_{j\in\mathcal{J}_0} s_j = 1$, the *identified utility set* associated with s is defined by

$$\tilde{\sigma}^{-1}(s) = \left\{\delta \in \mathbb{R}^{\mathcal{J}} : \tilde{\sigma}(\delta) = s\right\}. \tag{3.1}$$

Requiring non-zero market shares is a standard assumption in demand inversion of discrete-choice models; see, e.g., Lemma 1 of Berry and Haile (2014). A sufficient condition for this is that ε has a nowhere vanishing density on $\mathbb{R}^{\mathcal{J}_0}$ (see, e.g., Galichon and Hsieh (2019)), which is satisfied in many canonical ARUMs such as logit, probit, and mixed-logit, but may not be applicable in more general models.[7]

3.1. The Equivalence Theorem.

The Equivalence Theorem. Next we introduce a matching game between consumers and yogurts, which is essentially that of Demange and Gale (1985); our presentation of this model is inspired by the presentation in chapter 9 of Roth and Sotomayor (1992)[8]. In this matching model, one side of the market consists of a continuum of consumers, distinguished by type ε, while the other side is an equi-massed continuum of jars of yogurt, distinguished by alternative (brand) $j \in \mathcal{J}_0$. Let $\mathcal{M}(P, s)$ be the set of probability distributions on $\Omega \times \mathcal{J}_0$

[7]For example, in pure characteristic models (see, e.g., Song (2007)), one needs to further restrict the range of the mean utility and data to ensure nonzero shares. Moreover, we assume throughout that the demand model is correctly specified, so that the identified utility set in (3.1) is non-empty.

[8]Demange and Gale's model is discrete and extends the model of Shapley and Shubik (1971) beyond the transferable utility setting. See also Crawford and Knoer (1981), Kelso and Crawford (1982), Hatfield and Milgrom (2005). We formulate a slight variant here in that we (1) we allow for multiple agents per type and (2) do not allow for unmatched agents. However, this leaves analysis essentially unchanged.

with marginal distributions P and s; namely, $\pi \in \mathcal{M}(P, s)$ if and only if $\pi(B \times \mathcal{J}_0) = P(B)$ for all B (Borel-measurable subsets of Ω), and $\pi(\Omega \times \{j\}) = s_j$ for all $j \in \mathcal{J}_0$.

Let $f_{\varepsilon j}(u)$ be the transfer (positive or negative) needed by a consumer ε in order to reach utility level $u \in \mathbb{R}$ when matched with a yogurt j. Symmetrically, let $g_{\varepsilon j}(v)$ be the transfer needed by a yogurt j in order to reach utility level $v \in \mathbb{R}$ when matched with a consumer ε. (The connection between f and g and the primitives of the discrete choice model will be clarified below, in Eq. (3.2).) The functions $f_{\varepsilon j}(.)$ and $g_{\varepsilon j}(.)$ are assumed increasing for every ε and j. This matching game features *imperfectly transferable utility*; the case with perfectly transferable utility obtains when the f and g functions are identities, and are discussed separately in the next section.

Definition 3 (Equilibrium outcome)**.** An equilibrium outcome in the matching problem is an element (π, u, v), where π is a probability measure on $\Omega \times \mathcal{J}_0$, u and v are Borel-measurable functions on (Ω, P) and $(\mathcal{J}_0, s)$ respectively, such that:

(i) π has marginal distributions P and s: $\pi \in \mathcal{M}(P, s)$.

(ii) there is no blocking pair: $f_{\varepsilon j}(u_\varepsilon) + g_{\varepsilon j}(v_j) \geq 0$ for all $\varepsilon \in \Omega$ and $j \in \mathcal{J}_0$.

(iii) pairwise feasibility holds: if $(\varepsilon, j) \in Supp(\pi)$, then $f_{\varepsilon j}(u_\varepsilon) + g_{\varepsilon j}(v_j) = 0$.

We use u_ε (resp. v_j) to denote the value of u evaluated at $\varepsilon \in \Omega$ (resp. the value of v evaluated at $j \in \mathcal{J}_0$); in this definition, u_ε and v_j denote, respectively, the equilibrium payoffs of the consumer with utility shock ε and the yogurt of brand j. Condition (i) implies that if a random vector (ε, j) has distribution $\pi \in \mathcal{M}(P, s)$, then $\varepsilon \sim P$ and $j \sim s$. Hence, π is interpreted as the probability distribution that a consumer with utility shock ε matched with a yogurt of type j. In other words, $\pi(j|\varepsilon)$ denotes the conditional probability that an individual with utility shock ε chooses yogurt j, which is degenerate under Assumption 2. To understand condition (ii), consider that if there exists a consumer ε and a yogurt of type j for which $f_{\varepsilon j}(u_\varepsilon) + g_{\varepsilon j}(v_j) < 0$, then there exists $u' > u_\varepsilon$ and $v' > v_j$ such that $f_{\varepsilon j}(u') + g_{\varepsilon j}(v') = 0$. In other words, (u', v') are feasible for (ε, j) and strictly improve upon the equilibrium payoffs u_ε and v_j, which is ruled out in equilibrium. Condition (iii) implies that, if (ε, j) are actually matched, then their equilibrium payoffs u_ε and v_j should indeed be feasible–that is, the sum of the transfer to ε and the transfer to j should be zero.

The next theorem establishes that the demand inversion problem is equivalent to a matching problem. The proofs for this and all subsequent claims are in the appendix.

Theorem 1 (Equivalence theorem)**.** *Under Assumptions 1 and 2, consider a vector of market shares s that satisfies $s_j > 0$ and $\sum_{j\in\mathcal{J}_0} s_j = 1$. Consider a vector $\delta \in \mathbb{R}^{\mathcal{J}}$. Then, the two following statements are equivalent:*

(i) δ belongs to the identified utility set $\tilde{\sigma}^{-1}(s) = \left\{\delta \in \mathbb{R}^{\mathcal{J}} : \tilde{\sigma}(\delta) = s\right\}$ associated with the market shares s in the sense of Definition 2 in the discrete choice problem with $\varepsilon \sim P$;

(ii) there exists $\pi \in \mathcal{M}(P, s)$ and $u_\varepsilon = \max_{j\in\mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j)$ such that $(\pi, u, -\delta)$ is an equilibrium outcome in the sense of Definition 3 in the matching problem, where

$$f_{\varepsilon j}(u) = u \text{ and } g_{\varepsilon j}(-\delta) = -\mathcal{U}_{\varepsilon j}(\delta). \tag{3.2}$$

This theorem establishes an equivalence between demand inversion in a random utility model and the problem of finding equilibrium wages in a labor matching market, where each "firm" (corresponding to our consumers ε) only hires one "worker" (corresponding to our yogurts j). $-\delta_j$'s play the role of the salaries of the workers. An increase in δ_j (a decrease in $-\delta_j$) increases the utility of the consumers, just as a decrease in salary increases the profit of the firm.

The intuition behind this equivalence is that in a matching equilibrium, the transfers are adjusted so that everyone is happy with their own choices: consumer ε seeks the largest payoff she can obtain in a feasible union with a partner j demanding utility v_j. In the discrete choice model, ε seeks the largest utility she can get out of choosing an alternative j associated with systematic utility δ_j. This explains why we need to set $\delta_j = -v_j$: an increase in v_j makes yogurt j a less attractive option for any consumer ε, corresponding to a decrease in the systematic utility δ_j in the discrete choice model. Moreover, by plugging (3.2) into Definition 3, we see that: (i) the no-blocking condition is satisfied:

$$\forall\ \varepsilon,\ j: \quad u_\varepsilon = \max_{j'\in\mathcal{J}_0} \mathcal{U}_{\varepsilon j'}\left(\delta_{j'}\right) \geq \mathcal{U}_{\varepsilon j}(\delta_j); \tag{3.3}$$

and (ii) the feasibility pair condition is satisfied:

$$\text{if } (\varepsilon, j) \in Supp(\pi): \quad u_\varepsilon = \max_{j'\in\mathcal{J}_0} \mathcal{U}_{\varepsilon j'}\left(\delta_{j'}\right) = \mathcal{U}_{\varepsilon j}(\delta_j). \tag{3.4}$$

Our equivalence result states that the identified set of utilities in the discrete-choice demand problems corresponds to the equilibrium set of some matching problem. This matching equivalence result also raises the possibility of partial identification, or multiplicity of the identified set of utilities. Assumption 2 implies a unique demand map, and hence a unique allocation in the matching problem. However, just as in Shapley and Shubik

(1971), there may be multiple prices (corresponding to the utilities here) which support the equilibrium allocation. We will return to this below.

Without Assumption 2, the equilibrium in the associated matching game may no longer be unique. However, Theorem 1 continues to hold even without Assumption 2; that is, the equivalence results holds regardless of whether consumers' optimizing choices are unique or not. We discuss this in the appendix.

3.1.1. *"Yogurts choose consumers".* The details in Theorem 1 permit us to be more explicit about interpreting the yogurt demand problem as a two-sided market. In the conventional view ("consumers choose yogurts"), each consumer ϵ, given the utility vector δ, chooses a brand of yogurt to optimize $u_\epsilon \equiv \max_j \mathcal{U}_{\epsilon j}(\delta_j)$. This leads to the argmax mapping, $j^*(\epsilon)$, which, given Assumption 2, will be single-valued.

Analogously, from the "yogurts choose consumers" perspective, each yogurt j, given consumers' payoffs $\{u_\epsilon\}_\epsilon$, optimizes $\delta_j = \min_\epsilon \mathcal{U}_{\epsilon j}^{-1}(u_\epsilon)$. This leads to the argmax mapping $\epsilon^*(j)$ which, in typical cases (with a continuum of consumers and finite brands of yogurt) will be one-to-many onto (surjective).

Note that we have the property that any $\epsilon^+ \in \epsilon^*(j^+)$ implies $j^+ = j^*(\epsilon^+)$, and vice versa; intuitively, these two argmax mappings are, loosely speaking, "inverse" to each other. Thus, these two choice problems – "consumers choose yogurts" and "yogurts choose consumers" – satisfy *mutual assent*: any yogurt optimally chosen by a consumer must "assent" to be chosen by that consumer. An empirical consequence is that the market shares emerging from the two choice problems above are identical; the two problems are observationally equivalent.

3.2. **ARUM case.** When the random utility model is additive (ARUM) as in example 2.1, one has $f_{\varepsilon j}(u) = u$ and $g_{\varepsilon j}(-\delta) = -\mathcal{U}_{\varepsilon j}(\delta) = -\varepsilon_j - \delta_j$, so that the stability conditions become $u_\varepsilon + v_j \geq \varepsilon_j$ with equality for $(\varepsilon, j) \in Supp(\pi)$. As noted initially in Galichon and Salanié (2015), this problem is now equivalent to a matching problem with transferable utility, where the joint surplus of a match between a worker ε and a firm j is ε_j.

Generally, these are *optimal transport* (or Monge-Kantorovich) problems (see Galichon (2016)), and a well-known result in optimal transport theory states that the equilibrium matching under transferable utility maximizes the total surplus $\mathbb{E}\left[\varepsilon_{\tilde{j}}\right]$ over all the distributions of $\left(\varepsilon,\tilde{j}\right)$ such that $\varepsilon \sim P$ and $\tilde{j} \sim s$, that is π solves

$$\max_{\left(\varepsilon,\tilde{j}\right)\sim\pi\in\mathcal{M}(P,s)} \mathbb{E}_{\pi}\left[\varepsilon_{\tilde{j}}\right]. \tag{3.5}$$

This problem has a dual and, as it turns out, u and δ are solutions to the dual problem:

$$\begin{aligned} &\inf_{u,\delta:\delta_0=0}\left\{\mathbb{E}_P\left[u_\varepsilon\right]-\mathbb{E}_s\left[\delta_{\tilde{j}}\right]\right\} \\ s.t. \quad & u_\varepsilon-\delta_j\geq\varepsilon_j \;\forall\varepsilon\in\Omega,\; j\in\mathcal{J}_0. \end{aligned} \tag{3.6}$$

Hence, for ARUMs, the identified utility set for a market share vector s corresponds to the set of optimizers for the following finite-dimensional convex optimization problem:

$$\inf_{\delta:\delta_0=0}\left\{\mathbb{E}_P\left[\max_{j\in\mathcal{J}_0}\left\{\delta_j+\varepsilon_j\right\}\right]-\mathbb{E}_s\left[\delta_{\tilde{j}}\right]\right\}.$$

Below, in Section 4.2, we discuss algorithms for solving such problems, based on approximating the continuum market by a market populated by a large (but finite) number of consumers and jars of yogurt. In this case, the above problems become finite-dimensional linear programs coinciding with the assignment game analyzed in Shapley and Shubik (1971).

3.3. **Lattice structure of the identified utility set.** As a consequence of the Equivalence theorem, we apply a number of results from matching theory to describe basic properties of the identified set $\tilde{\sigma}^{-1}(s)$. Next, we show that the set-valued function $s\to\tilde{\sigma}^{-1}(s)$ is isotone[9] (in a sense to be made precise) and that $\tilde{\sigma}^{-1}(s)$ has a lattice structure.

The literature on the estimation of discrete choice models has favored an approach based on imposing conditions guaranteeing invertibility of demand, or equivalently situations in which $\tilde{\sigma}^{-1}(s)$ is restricted to a single point. In particular, Berry, Gandhi, and Haile (2013) (hereafter BGH) provide conditions under which $\tilde{\sigma}^{-1}(s)$ should contain at most one point, from which it also follows that the map $s\to\tilde{\sigma}^{-1}(s)$ is isotone on its domain. In contrast, our approach here imposes minimal assumptions, and one must consider *non-invertible* models, in which the demand map (1.2) is not one-to-one and $\tilde{\sigma}^{-1}(s)$ is a set. In this case, we need to generalize the notion of isotonicity which applies to the identified set $\tilde{\sigma}^{-1}(.)$. The next theorem states that the correct generalization is the notion of isotonicity with respect

[9] Isotone has the meaning of (monotone) increasing, and is a standard term used in the literature on lattices (see Topkis (1998) for a reference on lattices and isotonicity in economics).

to *Veinott's strong set order*.[10] For the following, recall the lattice "join" and "meet" operators ($\wedge$ and $\vee$) are defined by $(\delta \wedge \delta')_j := \min\left\{\delta_j, \delta'_j\right\}$ (componentwise minimum) and $(\delta \vee \delta')_j := \max\left\{\delta_j, \delta'_j\right\}$ (componentwise maximum).

Theorem 2. *The set-valued function $s \to \tilde{\sigma}^{-1}(s)$ is isotone in Veinott's strong set order, i.e. if $\delta \in \tilde{\sigma}^{-1}(s)$ and $\delta' \in \tilde{\sigma}^{-1}(s')$ with $s \leq s'$, then $\delta \wedge \delta' \in \tilde{\sigma}^{-1}(s)$ and $\delta \vee \delta' \in \tilde{\sigma}^{-1}(s')$.*

In the special case where $s \to \tilde{\sigma}^{-1}(s)$ is a singleton, we recover the isotonicity of the inverse demand as in BGH. Going further, by taking $s = s'$ in Theorem 2, we obtain that, whenever it is non-empty, the set $\tilde{\sigma}^{-1}(s)$ is a *lattice*[11]:

Corollary 1. *Under Assumption 1 and 2, if $\tilde{\sigma}^{-1}(s)$ is non-empty, it is a lattice. That is, if $\delta, \delta' \in \tilde{\sigma}^{-1}(s)$, then both $(\delta \wedge \delta'), (\delta \vee \delta') \in \tilde{\sigma}^{-1}(s)$.*

This result is well-known in matching theory since Demange and Gale (1985), who showed that the set of payoffs which ensures a stable allocation is a lattice whenever it is non-empty. This implies that the set of identified utilities has a "maximal" (resp. "minimal") element which is composed of the *component-wise* upper- (resp. lower-) bounds among all the utility vectors in the identified set. The upper bound corresponds to the unanimously most preferred stable allocation for the consumers ("consumer-optimal") and the unanimously least preferred stable allocation for the yogurts; conversely, the lower bound corresponds to the unanimously most preferred stable allocation for the yogurts ("yogurt-optimal") and least preferred for the consumers. Formally, define

$$\tilde{\delta}_j^{\min}(s) = \min\left\{\delta_j : \delta \in \tilde{\sigma}^{-1}(s)\right\} \text{ and } \tilde{\delta}_j^{\max}(s) = \max\left\{\delta_j : \delta \in \tilde{\sigma}^{-1}(s)\right\}.$$

Then the lattice property implies:

(i) The set $\tilde{\sigma}^{-1}(s)$ has a minimal and a maximal element:

$$\tilde{\delta}^{\min}(s) \in \tilde{\sigma}^{-1}(s) \text{ and } \tilde{\delta}^{\max}(s) \in \tilde{\sigma}^{-1}(s).$$

(ii) Any $\delta \in \tilde{\sigma}^{-1}(s)$ is such that

$$\tilde{\delta}^{\min}(s) \leq \delta \leq \tilde{\delta}^{\max}(s).$$

[10]See e.g. Veinott (2005). Veinott's strong set order provides an ordering over sets. Let $\mathcal{X}$ and $\mathcal{X}'$ be two subsets in $\mathbb{R}^d$; we say that $\mathcal{X} < \mathcal{X}'$ in the Veinott strong set order iff $\forall x \in \mathcal{X}, x' \in \mathcal{X}'$, the "join" (or componentwise minimum) $x \wedge x' \in \mathcal{X}$ and the "meet" (or componentwise maximum) $x \vee x' \in \mathcal{X}'$.

[11]Whether the identified set is empty is considered in the Section 7.

(iii) $\tilde{\sigma}^{-1}(s)$ is point-identified if and only if

$$\tilde{\delta}^{\min}(s) = \tilde{\delta}^{\max}(s).$$

Practically, most applications of partially identified models focus on computing the component-wise upper and lower bounds of the identified set of parameters; for general partially identified models, the vector of component-wise bounds will typically lie *outside* the (joint) identified set of parameters. In contrast, our lattice result here implies that these component-wise upper and lower bounds constitute *sharp* upper and lower bounds for the parameter vector as a whole, in the sense that they are attainable for selection mechanisms which place all probability on the highest (for upper bound) or lowest (for lower bound) payoffs for consumers.

In addition, the matching literature provides algorithms to compute these extremal elements, which can be directly used to assess partial identification: indeed, $\tilde{\sigma}^{-1}(s)$ is a single element (point-identified) if and only if its minimal and maximal elements coincide. This flexibility in handling models for which the researcher may not know *a priori* whether model parameters are point- or partially-identified is an important contribution of the matching approach developed in this paper. We turn to these algorithms next.

4. Matching-based Algorithms

The equivalence established in Theorem 1 between matching and discrete-choice models allows us to leverage several matching algorithms for both NARUMs and ARUMs. The use of these algorithms in the empirical discrete-choice literature is new. In addition, matching-based algorithms have several advantages over existing procedures: (i) they can handle non-invertibility of the demand map, as all of these algorithms allow for the case of partial identification, when multiple utility vectors can rationalize the same set of market shares; and (ii) these algorithms do not require smoothness of the demand map and therefore can handle some well-known models – like the pure characteristics model – which have non-smooth demand maps. In contrast, existing algorithms for demand inversion often rely on directly solving the demand map $s_j = \sigma_j(\delta)$ for δ using fixed-point iterations or nonlinear-equation solvers, which typically requires smoothness of the demand map, and also rules out non-invertibility of the demand map.

We introduce three matching-based algorithms in this section. The first is an algorithm for matching models with imperfectly transferable utility (ITU) which can be used for demand inversion in both ARUMs or NARUMs. This algorithm, called *Market Share Adjustment*, is essentially an "accelerated" version of the classic deferred acceptance algorithm (Gale and Shapley (1962), Crawford and Knoer (1981)), which rely on iteratively adjusting the payoffs of the potential partners to achieve equilibrium. The second and third algorithms, in Section 4.2, are methods for computing stable allocations in two-sided matching models with transferable utility (TU), and apply only to ARUMs. We consider a *linear-programming* approach based on the classic Shapley and Shubik (1971) assignment game, and a version of the *auction algorithm* of Bertsekas (1992), augmented to produce bounds for partially-identified settings.

While the model in this paper assumes a continuum of agents on each side of the market, for computational purposes we approximate this with a finite market populated by an equal (and large but finite) number, denoted N, of consumers and jars of yogurt.[12] On the consumer side, each consumer $i \in \{1, ..., N\}$ is characterized by a value of the utility shock ε_i drawn i.i.d. (across i) from P, the distribution of the utility shocks. For a given vector of market shares $(s_0, s_1, \ldots, s_J)$, the number of jars of each brand j of yogurt are set proportionately to the observed market share; that is, $m_j \approx N s_j \in \mathbb{N}$ of yogurts of type j and, if needed, m_j has been rounded to an adjacent integer so that $\sum_{j \in \mathcal{J}_0} m_j = N$. Throughout we maintain the utility normalization $\delta_0 = 0$.

4.1. Deferred-acceptance type algorithm (for both NARUMs and ARUMs).

Theorem 1 establishes an equivalence between the identified utility set and equilibrium payoffs in a two-sided matching game with imperfectly transferable utility. Hence, for computing the identified utilities one could use the deferred-acceptance algorithms developed in Crawford and Knoer (1981) and Kelso and Crawford (1982) which are generalizations of Gale and Shapley's (1962) deferred-acceptance algorithm. But these algorithms are very slow and inefficient, especially in the common situation where there are fewer products (i.e. "brands of yogurt") than consumers.

4.1.1. *Market shares adjusting algorithm (MSA).* As with all deferred-acceptance algorithms, there are two versions of the algorithm – the "consumer-proposing" and "yogurt-proposing" versions – return (resp.) the lattice upper bound or lower bound on the utility parameters. In the case when the model is point identified, the upper bound and lower

[12] In Appendix B we present the *semi-discrete* algorithm for ARUMs, which is exact in that it computes the continuum problem directly. However, as we explain there, its use is limited to particular parametrizations of the random shocks, and also does not handle the partial-identified case.

bound will coincide; hence running both versions of this algorithm yields a data-driven assessment of whether the model is point- or partially-identified.

In the "consumer proposing" version, the utilities (δ's) start at a high level, and consumers choose, in successive rounds, jars of yogurts which maximize their utilities. Between rounds, the systematic utilities pertaining to the brands of yogurts in excess demand (ie. chosen by more consumers than available jars) are decreased by an adjustment factor. Bidding continues until a round is reached where the reference brand of yogurt ($j = 0$) is in excess demand: that is, when the number of consumers choosing jars of brand 0 is greater or equal to the number of its available jars. In the original Kelso and Crawford (1982) version, the adjustment is done for *each jar* of yogurt separately, leading to very slow convergence with a large number of consumers and jars. The MSA algorithm speeds this up by adjusting the utilities for *all jars of the same brand* of yogurt simultaneously. Since this accelerated process can lead to "overshooting" (in which utilities move below their equilibrium values), the MSA algorithm involves running a deferred-acceptance procedure multiple times with successively smaller adjustment factors.[13]

We present below the pseudo code for the consumer-proposing version of the MSA algorithm, which obtains the lattice upper bound; Appendix A.5 contains a version of the algorithm which yields the lattice lower bound. Define $\bar{\delta}_j = \sup_{i \in \{1,\ldots,N\}} \mathcal{U}^{-1}_{\varepsilon_{ij}} \left(\mathcal{U}_{i0}\left(\delta_0\right)\right)$. Clearly, $\bar{\delta}$ is an upper bound for the stable payoffs and for the lattice upper bound. Let η^{tol} denote a small adjustment factor, which is a design parameter for the algorithm.

Algorithm 1 (Consumer-proposing MSA)**.**
Define $\delta_j^{init} = \sup_{i \in \{1,\ldots,N\}} \mathcal{U}^{-1}_{\varepsilon_{ij}} \left(\mathcal{U}_{i0}\left(\delta_0\right)\right)$ *# Starting values above upper-bound*

Begin **Adjustment (outer) Loop**

Initialize $\eta^{init} >> \eta^{tol}$ *and* $\delta_j^{init} = \delta_j^{return}$.

Repeat:

Call **Deferred-Acceptance Loop** *with* $\left(\delta_j^{init}, \eta^{init}\right)$ *which returns* (δ', η')

Set $\delta^{init} \leftarrow \delta' + 2\eta'$ *and* $\eta^{init} \leftarrow \eta'$

Until $\delta_j^{return} < \delta_j^{init}$ *for all* $j \in \mathcal{J}$. *# Algo stops if all* δ *have decreased*

---

[13]This adjustment factor plays a role analogous to the step size in optimization procedures. One typically chooses a larger step size in initial *exploration* phases to move parameters away from regions where the optimum is unlikely to be. Choosing a larger step size speeds up the routine, but if the step size is too large, one might miss ("overshoot") the optimum. Therefore, in later *exploitation* phases, one decreases the step size to achieve a better accuracy. Similar heuristics are used also to set the temperature parameter in simulated annealing, or the "learning rate" in machine learning procedures.

End **Adjustment Loop**

Begin **Deferred-acceptance (inner) Loop**

Require $\left(\delta_j^{init}, \eta^{init}\right)$.

Set $\eta = \eta^{init}$ *and* $\delta = \delta^{init}$.

While $\eta \geq \eta^{tol}$ # *Run as long as tol. factor above threshold* η^{tol}

If $j \in \arg\max_j \mathcal{U}_{\varepsilon j}(\delta_j)$ *then* $\pi_{ij} = 1$ *else* $\pi_{ij} = 0$ # *i is matched to optimizing j*

If $\sum_i \pi_{i0} < m_0$, *then for all* $j \in \mathcal{J}$ # *If brand 0 in excess supply then*

if $\sum_i \pi_{ij} > m_j$ *then* $\delta_j \leftarrow \delta_j - \eta$. # *decrease* δ_j, *for brands j w/ excess demand*

Else # *Else if brand 0 in excess demand (overshooting)*

$\delta_j \leftarrow \delta_j + 2\eta$ *for all* $j \in \mathcal{J}$ # *Increase all* δ *(except* δ_0*)*

$\eta \leftarrow \eta/4$. # *Decrease the tol. factor for next loop*

End While

Return $\delta^{return} = \delta$ *and* $\eta^{return} = \eta$.

End **Deferred-acceptance Loop**

As shown above, the consumer-proposing MSA consists of a deferred-acceptance loop nested inside of an adjustment loop. In the deferred-acceptance loop, two cases can occur. In the "good case", the deferred-acceptance loop starts with values $\delta_{j\in\mathcal{J}}$ above the lattice upper bound and an adjusting factor η small enough, then all $\delta_{j\in\mathcal{J}}$ will reach the lattice upper bound without overshooting. In the "bad case", that is when the deferred-acceptance loop starts with one $\delta_{j\in\mathcal{J}}$ below its upper bound and an adjusting factor η small enough, then this δ_j will not decrease during the loop. The outer adjustment loop repeatedly calls the deferred-acceptance loop for decreasing values of the increment η. It terminates when the utilities outputted by the deferred-acceptance loop have all decreased (indicating that the approximation loop is in the "good case"); otherwise, the utilities are increased and the deferred-acceptance loop is called again.

While we have not yet formally proven convergence of the MSA algorithm, we find that it terminates remarkably quickly in all our simulations, relative to the Crawford and Knoer (1981) algorithm. In practice, one can assess the convergence of the algorithm in the outer loop by comparing the actual market shares with the predicted market shares evaluated using the δ's returned in each call to the inner loop.

4.2. **Transferable Utility Matching algorithms (for ARUMs).** For ARUMs, as discussed in Section 3.2 above, we can use algorithms for matching models with transferable utility (TU). These are equivalent to optimal transport problems (e.g., Galichon (2016)), a class of convex programs which can be solved efficiently by linear programming (LP) or auction algorithms.[14] We contribute two novel modifications to the literature. For LP, we introduce a formulation that *simultaneously* inverts multiple demand maps. For the auction algorithm, we augment the classical algorithm so it can produce bounds for the partially-identified models.

4.2.1. *Linear programming (Shapley-Shubik).* For ARUMs, the solution can be obtained from the linear program (3.6). Specifically, for the large but finite discretization described above, this linear program is equivalent to the dual of the Shapley and Shubik (1971) assignment game (cf. Eq. (3.6)):

$$\begin{aligned} \inf_{u_i,\delta_j} \quad & \sum_{i=1}^{N} \frac{1}{N} u_i - \sum_{j\in\mathcal{J}_0} s_j \delta_j \\ s.t. \quad & u_i - \delta_j \geq \varepsilon_j^i, \quad \forall i, j. \end{aligned} \tag{4.1}$$

Shapley and Shubik (1971) show that the set of optimizers in (4.1) is a lattice, with bounds equal to the consumer-optimal and yogurt-optimal payoffs, but they do not discuss the computation of these bounds. In principle, the upper (resp. lower) bounds can be obtained from the following problem:

$$\max_{u,\delta,\pi} \text{ (resp. } \min_{u,\delta,\pi}) \sum_{j=0}^{J} \delta_j \quad \text{s.t. } (u,\delta) \in \{\text{arginf (4.1)}\}.$$

This is a "bilevel" program, as the solutions to the LP in (4.1) are used as the inputs into a second LP.[15] As is well-known, we can collapse a bilevel LP into a regular LP by replacing the lower-level LP (corresponding to (4.1)) with its optimality conditions. That is, the

[14] In Appendix B, we also describe the *semi-discrete* algorithm—a cutting-edge optimal transport solution approach. However, it requires the unobserved taste vector to be (jointly) uniformly distributed over a polyhedron, which limits its general application.

[15] See Dempe, Kalasshnikov, Pérez-Valdés, and Kalashnykova (2015).

upper (resp. lower) bounds can be obtained from the following LP:

$$\begin{aligned}
\max_{u,\delta,\pi} \text{ (resp. } \min_{u,\delta,\pi}) \quad & \textstyle\sum_{j=0}^{J} \delta_j \\
\text{s.t.} \quad & \textstyle\sum_{i=1}^{N} \pi_{ij} = s_j, \quad \forall j \\
& \textstyle\sum_{j=0}^{J} \pi_{ij} = \frac{1}{N}, \quad \forall i \\
& \pi_{ij} \geq 0, \quad \forall i, j \\
& u_i - \delta_j \geq \epsilon_{ij}, \quad \forall i, j \\
& \textstyle\sum_{i=1}^{N} \sum_{j=0}^{J} \pi_{ij}\epsilon_{ij} = \frac{1}{N} \sum_{i=1}^{N} u_i - \sum_{j=0}^{J} s_j \delta_j, \\
& \delta_0 = 0.
\end{aligned} \tag{4.2}$$

Constraints 1-3 in (4.2) are the constraints of the primal assignment game (3.5), whereas constraint 4 appears in the dual problem (3.6). Constraint 5 equates the primal and dual objectives at the optimum. Taken together, these 5 constraints characterize the optimizing (u, δ) from (3.6).[16] Constraint 6 is our maintained normalization.

Combining LP problems. A further benefit of the LP approach is that multiple demand inversion problems for different markets can be *combined* and solved simultaneously. Specifically, suppose there are $t = 1, \ldots, T$ markets. Instead of solving Problem (4.1) for each of the T markets separately, we combine these T problems into *one* problem to invert *all* demand maps *simultaneously*:

$$\begin{aligned}
\inf_{u_{ti},\delta_{tj}} \quad & \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \frac{1}{N} u_{ti} - \sum_{j \in \mathcal{J}_0} s_{tj} \delta_{tj} \right) \\
s.t. \quad & u_{ti} - \delta_{tj} \geq \varepsilon_{tj}^{i}, \quad \forall i, j, t
\end{aligned} \tag{4.3}$$

where all of the subscripts are augmented by the market index t. Since the decision variables (u_{ti}, δ_{tj}) and the associated constraints are market-specific, the resulting constraint coefficient matrix has a block-diagonal structure, which enables the use of efficient parallel sparse matrix routines for modern LP solvers. We utilize this simultaneous demand inversion approach in the empirical application below, and confirm how one large but sparse problem (involving a larger number of parameters and constraints) is solved much more quickly than T small problems.

[16]See, for instance, Mangasarian (1969).

4.2.2. *Auction algorithms.* Auction-type algorithms à la Bertsekas (1992) provide an alternative approach to linear programming methods for solving TU-matching models. In these algorithms, unassigned persons bid simultaneously for objects, decreasing their systematic utilities (or equivalently raising their prices). Once all bids are in, objects are assigned to the highest bidder. The procedure is iterated until no one is unassigned. The description here follows Bertsekas and Castanon (1989). We let $\kappa \in \{1, \ldots, N\}$ index jars of yogurt, where $j(\kappa) \in \mathcal{J}_0$ denotes the brand identity for the κ-th jar of yogurt.

We define the prices p_κ as negative systematic utilities: $p_\kappa = -\delta_\kappa$.

Algorithm 2 (Auction)**.** *Start with an empty assignment and a given vector of prices p_κ and set a scale parameter $\eta > 0$.*

Bidding phase

(a) Each currently unassigned consumer i chooses the jar $\kappa^\star$ to maximize utility:

$$\mathcal{U}_{\varepsilon_i \kappa^\star}(-p_{\kappa^\star}) = \max_\kappa \mathcal{U}_{\varepsilon_i \kappa}(-p_\kappa). \tag{4.4}$$

(b) Consumer i's bid is set to:

$$b_{i\kappa^\star} = p_{\kappa^\star} + \mathcal{U}_{\varepsilon_i \kappa^\star}(-p_{\kappa^\star}) - w_i + \eta \tag{4.5}$$

where w_i denotes the utility from consumer i's second-best choice:

$$w_i = \max_{j(\kappa) \neq j(\kappa^\star)} \mathcal{U}_{\varepsilon_i \kappa}(-p_\kappa) \tag{4.6}$$

Assignment phase

Jar κ is assigned to its highest bidder $i^\star$ its price is raised to bidder $i^\star$'s bid:

$$p_\kappa := b_{i^\star \kappa} \tag{4.7}$$

Final step

When no one is left unassigned the solution δ_j is recovered as:

$$\delta_j = -\min_{\kappa \in j(\kappa)} p_\kappa \tag{4.8}$$

Intuitively, the algorithm implements a Walrasian-style bidding procedure. In each round, each unassigned consumer bids for his favorite jar of yogurt. His bid (Eq. (4.5)) is equal to the difference between the utilities from his most-preferred and second-most-preferred brands of yogurt (plus an extra $\eta > 0$ factor to ensure that prices are increasing each round). The consumer which makes the highest bid for a jar is assigned to it, and its price is increased by the amount of the bid; a consumer previously assigned to this jar becomes unassigned and bids in the next round. The algorithms stops when all individuals are assigned. The performance of the algorithm is considerably improved by applying it several times, starting with a large value of η and gradually decreasing it.

Computing utility bounds. The auction algorithm as described above works for the case when the demand map is invertible (so that the identified set of utilities is a singleton). In practice, we do not know *a priori* whether the demand map is invertible or not; hence, we must extend the algorithm to allow for partial identification. To do this, we note that the auction algorithm described above provides the optimal *allocation* of consumers to jars of yogurt. The bounds on the equilibrium utilities δ's which support this optimal allocation can be computed using the *Bellman-Ford algorithm*, which we describe here.

From the auction algorithm, we know the optimal allocation which we denote by a set of indicator variables π_{ij} (for $i = 1, \ldots, N$ and $j \in \mathcal{J}_0$) which is equal to one if consumer i matches with a jar of yogurt brand j, and zero otherwise. We consider the dual problem where we look for equilibrium payoffs to consumers (denoted u_i for $i = 1, \ldots, N$) and to brands of yogurt (corresponding to δ_j for $j \in \mathcal{J}_0$), with the normalization $\delta_0 = 0$.

We can solve for the minimal (lower-bound) equilibrium payoffs $\{u_i, \delta_j\}_{i,j}$ as a fixed point of the mapping

$$u_i = \max(u_i, \max_{j \in \mathcal{J}_0}(\mathcal{U}_{\varepsilon_{ij}}(\delta_j))), \quad \delta_j = \max(\delta_j, \max_{i:\pi_{ij}>0} \mathcal{U}^{-1}_{\varepsilon_{ij}}(u_i)), \quad \delta_0 = 0 \tag{4.9}$$

which is an isotone operator, given Assumption 1(b). A fixed point of (4.9) satisfies

$$u_i \geq \mathcal{U}_{\varepsilon_{ij}}(\delta_j) \ \forall i, j; \quad u_i \leq \mathcal{U}_{\varepsilon_{ij}}(\delta_j) \ \forall i, j : \pi_{ij} > 0$$

or, rearranging, we get $u_i \geq \mathcal{U}_{\varepsilon_{ij}}(\delta_j) \ \forall i, j$ with equality for $\pi_{ij} > 0$, which corresponds to the "no blocking pair" condition, cf. Eqs. (3.3,3.4). Hence, the lower-bound payoffs can be computed by iterating on (4.9) from the initial values $\{u_i = -\infty; \ \delta_j = -\infty, \ j \neq 0; \ \delta_0 = 0\}$.

Analogously, starting with values of $+\infty$ and iterating on the following operator returns upper bounds:

$$\delta_j = \min(\delta_j, \min_{i \in \mathcal{I}} \mathcal{U}^{-1}_{\varepsilon_{ij}}(u_i)), \quad u_i = \min(u_i, \min_{j:\pi_{ij}>0}(\mathcal{U}_{\varepsilon_{ij}}(\delta_j))), \quad \delta_0 = 0. \tag{4.10}$$

4.3. **Implementation.** We have developed R packages containing fast and efficient implementations of the algorithms described in this paper. They are designed to enable user-friendly access for researchers to popular matching methods, and are used in the next section to benchmark the relative performance of each algorithm. Bonnet, Galichon, Hsieh, O'Hara, and Shum (2018a) collects several auction and linear programming algorithms into an R package, utilizing C++ code provided by Walsh and Dieci (2017). Finally, a parallelized implementation of the MSA is contained in a separate package Bonnet, Galichon, Hsieh, O'Hara, and Shum (2018b). Links to the packages and installation instructions are provided in the bibliography references.

5. Numerical experiments

We test our algorithms on two different models: the first one is the additive pure characteristics model, and the second one is a "two-store" mixture version of the vertical differentiation model which is *not* invertible, so that multiple values of utilities are consistent with a set of market shares.

5.1. **Example 1: the pure characteristics model.** The literature has emphasized several reasons to prefer the pure characteristics model to the random coefficient logit model.[17] However, it has not been often utilized in empirical work because the non-smooth demand map of this model imposes computational hurdles.

Here we consider the performance of our matching-based algorithms in computing the pure-characteristics model. Our algorithm, which we call "Matching-NFXP", is essentially the BLP algorithm (Berry, Levinsohn, and Pakes (1995)), except that the inner loop is replaced by one of the matching algorithms in Section 4. For the first set of simulations, we implement the combined LP defined in Eq. (4.3) for the inner loop.

We consider the following specification: The utility of consumer i who chooses product j in market m is generated by:

$$u_{mij} = \beta_0 - \beta_p p_{mj} + \beta_1 x_{mj1} + \beta_2 x_{mj2} + \beta_3 x_{mj3} + \xi_{mj}, \tag{5.1}$$

where $x_{mjk}, k = 1, 2, 3$ are observed, exogenous product attributes, and p_{mj} is the price of product j in market m that is correlated with the unobserved product attribute ξ_{mj}. $(\beta_p, \beta_1, \beta_2, \beta_3)$ are assumed to be individual-specific random coefficients. There are 100 markets and 5 products (including the outside goods). To generate the market shares, we simulated 1000 consumers for each market.

[17] Models with logit errors have properties that may be undesirable for welfare analysis: They restrain substitution patterns and utility grows without bounds as the number of products in the market grows. See Berry and Pakes (2007) and Ackerberg and Rysman (2005)

To benchmark performance, we compare our "Matching-NFXP" approach to two alternatives suggested in the literature. The first, denoted "BLP-MPEC", comes from Berry and Pakes (2007), who suggest smoothing out the non-smooth pure-characteristics demand map by adding small logit errors to each alternative, and then using the BLP estimation procedure. In our simulations, we utilize the MPEC (Dubé, Fox, and Su 2012) version of BLP. The second alternative, denoted "PSL", is the algorithm proposed in Pang, Su, and Lee (2015). The details of the data generating process and three algorithms can be found in Appendix C.

In Table 1, we report the RMSE, bias, proportion of runs converged, and the average runtime across 20 Monte Carlo repetitions. We consider two model specifications: In Model I, we estimate the location parameters and fix all scale parameters. In Model II, we estimate the location parameter and the scale parameter associated with the endogenous price, fixing the rest of the scale parameters.

Table 1. Numerical Performances: Estimating the Demand Parameters in Pure Characteristics Models

			RMSE					Bias				
algorithm	convergence(%)	runtime (sec.)	cons	σ_p	x_1	x_2	x_3	cons	σ_p	x_1	x_2	x_3
Model I												
Matching-NFXP	100	3.09	0.08		0.07	0.07	0.06	-0.03		-0.01	-0.02	0.01
PSL	90	409.65	0.08		0.07	0.07	0.06	0.01		-0.01	-0.02	0.01
BLP-MPEC	100	79.38	0.27		0.16	0.17	0.16	0.14		0.02	0.01	-0.09
Model II												
Matching-NFXP	100	120.19	0.09	0.07	0.07	0.07	0.07	-0.02	0.00	-0.01	-0.01	0.01
PSL	85	834.48	0.14	0.35	0.09	0.10	0.11	-0.09	-0.28	0.06	0.05	0.08
BLP-MPEC	100	260.03	0.26	0.13	0.17	0.17	0.16	0.12	-0.06	0.04	0.03	-0.07

Note: The numbers are averages across 20 Monte Carlo repetitions. There are 100 markets and 5 products. We simulated 1000 consumers for each market. cons, x_1, x_2, x_3 are the location parameters, whereas σ_p is the scale parameter of price. Following Berry and Pakes (2007), the location parameter of price is normalized to -1.

We first discuss the numerical accuracy. In Model I, Matching-NFXP and PSL have similar RMSE, which is roughly half of that of BLP-MPEC. In Model II, Matching-NFXP clearly dominates the other two alternatives. In particular, Matching-NFXP delivers nearly an unbiased estimate for the standard deviation of the random coefficient of price (σ_p), while the other two approaches falter. In terms of computing speed, our method is on average

130 times faster than PSL and 26 times faster than BLP-MPEC in Model I. In model II, our method outperforms PSL by a factor of 7 and outperform BLP-MPEC by a factor of 2. These simulations demonstrate the superior performance of the matching-based approach for estimating the pure characteristics model.

TABLE 2. Demand Inversion in Pure Characteristics Models: comparing different algorithms

Algorithms	Draws	Brands	RMSE	CPU time
BLP contract. map.	1,000	5	0.070	0.032
LP	1,000	5	0.029	0.083
Auction	1,000	5	0.029	0.010
MSA	1,000	5	0.029	18.776
BLP contract. map.	1,000	50	0.045	0.283
LP	1,000	50	0.013	0.313
Auction	1,000	50	0.013	0.046
BLP contract. map.	1,000	500	0.018	2.774
LP	1,000	500	0.004	3.869
Auction	1,000	500	0.005	0.426
BLP contract. map.	10,000	5	0.072	0.331
LP	10,000	5	0.014	0.304
Auction	10,000	5	0.014	0.117
MSA	10,000	5	0.014	2.608
BLP contract. map.	10,000	50	0.044	2.890
LP	10,000	50	0.006	4.446
Auction	10,000	50	0.006	0.659
BLP contract. map.	10,000	500	0.017	38.519
LP	10,000	500	0.002	64.061
Auction	10,000	500	0.002	5.185

Note: The numbers are averaged from 50 Monte-Carlo replications. Demand inversion for the pure characteristics model with 5, 50 and 500 brands of yogurt and $1,000$ and $10,000$ draws of taste shocks. The column "RMSE" corresponds to the root mean squared errors of the estimated δ_j.

Since our matching algorithms are used only in the inner loop of the "Matching-NFXP" procedure, we next focus on the demand inversion step, and compare the performance of our matching-based algorithms to alternative existing approaches. Table 2 summarizes the numerical performance of three matching-based algorithms: (i) LP, (ii) Auction, and (iii) MSA; the (iv) BLP contraction mapping (which adds logit errors to the utilities of each choice to smooth the demand map) is also included as a benchmark.[18]

The numerical accuracy is almost identical across all three matching algorithms (LP, Auction, MSA). In comparison, the RMSE of the BLP contraction mapping is about three times larger. This arises from the additional approximation error due to the introduction of additive logit errors to smooth the mapping. For computational speed, the auction algorithm is the fastest by a wide margin: even under the most demanding scenario (500 brands and 10,000 simulated consumers), it only takes 5 seconds, which far outstrips other approaches.

On the other hand, while MSA is the slowest algorithm, and does not scale up well with the number of brands, it is a general-purpose algorithm that also applies to NARUMs.[19] It is therefore not surprising that MSA is not as fast as the other methods since it does not exploit the additive separability in ARUMs. In the next example, we spotlight the performance of the MSA algorithm in a partially-identified NARUM.

5.2. **Example 2: a partially-identified model.** An important contribution of our approach is that we can handle models which are not invertible, and hence the utilities may not be point identified. We consider an example which highlights this.

There are three goods $y = 1, 2, 3$, and the unknown parameters are the quality of each good are $\delta_1, \delta_2, \delta_3$, with the normalization $\delta_1 = 0 < \delta_2, \delta_3$. Supply is generated by two stores: in store 1, prices are $p_1^1 < p_2^1 < p_3^1$, while in store 2, prices are $p_1^2 \leq p_3^2 < p_2^2$.

Consumers are heterogeneous in their willingness-to-pay for quality, given by $\theta \sim U[0,1]$. Each consumer has a 1/2 chance of going to either store. Hence, in this model, the consumer-idiosyncratic shocks ε include two components: the heterogeneity θ as well as the prices that they face. Consumers' utilities are given by $\mathcal{U}_{\varepsilon,y}(\delta_y) = \theta\delta_y - p_y$. Let s_y^j denote the (unobserved) market share of good y at store $j = 1, 2$. The observed market shares are mixtures of market shares at the two stores:

$$s_y = 0.5(s_y^1 + s_y^2), \quad y = 1, 2, 3. \tag{5.2}$$

[18]Complete details on these simulations are provided in Appendix C.

[19]Interestingly, with 5 brands, MSA is faster with 10,000 draws than with 1,000: this may appear counter-intuitive, but increasing the number of draws improves accuracy during the loop and therefore reduces total computation times.

Consider goods y and y' with $\delta_y \geq \delta_{y'}$. Consumer i chooses good y over y' at store j iff[20]

$$\mathcal{U}_{\varepsilon_i,y}(\delta_y) \geq \mathcal{U}_{\varepsilon_i,y'}(\delta_{y'}) \Leftrightarrow \theta \geq \hat{\theta}_j(y,y') \equiv \min\left\{1, \max\left\{0, \frac{p_y^j - p_{y'}^j}{\delta_y - \delta_{y'}}\right\}\right\} \tag{5.3}$$

In this model, the δ's can be partially identified, and for the simulation exercise, we consider such a case. Let $p_1^1 = 1$, $p_2^1 = 2$, and $p_3^1 = 3$ be the prices at store 1, and $p_1^2 = 1$, $p_2^2 = 2$, and $p_3^2 = 1$ at store 2. The observed market shares are given by $(s_1, s_2, s_3) = (0.25, 0.25, 0.5)$. Given $\delta_1 = 0$, the quality parameters (δ_2, δ_3) that rationalize these market shares is:

$$C = \{(\delta_2, \delta_3) : \delta_2 = 2, \delta_3 \in [1,3]\} \tag{5.4}$$

As we expect, we is a lattice, with minimal vector $(0, 2, 1)$ and maximal vector $(0, 2, 3)$.

Computational results using the market shares adjustment (MSA) algorithm, are given in Table 3. The algorithm performs as expected, and running the two versions of the MSA algorithm (described in Section 4.1.1 and Appendix A.5) accurately recovers the upper and lower bounds.

Table 3. Mixture vertical differentiation: 3 goods

	True	$N = 1000$		$N = 100$	
	δ_y	Upper-bound	Lower-bound	Upper-bound	Lower-bound
δ_2	2	2.004	1.996	2.041	1.961
δ_3	[1,3]	3.005	0.995	3.051	0.951

We normalized $\delta_1 = 0$. The aggregate market shares are $(s_1, s_2, s_3) = (0.25, 0.25, 0.5)$, and prices: store 1 $(1, 2, 3)$, store 2 $(1, 2, 1)$. Lower and upper bounds are computed with the MSA algorithm. N denotes the number of consumers and jars of yogurts used in the MSA algorithm.

To continue, we consider an expanded version of this model involving 5 stores, and 8 goods. The prices and market shares for this example are given in Table 4. Since, these market shares were chosen arbitrarily, we do not know *a priori* whether the utilities vectors δ are point- or partially-identified.[21] Hence we run both the consumer- and yogurt-proposing

[20]Note that, according to (5.3), every consumer prefers the cheapest good between two same-quality goods. If two goods have the same price and quality, consumers are indifferent between them, so any demand is rationalizable. Our numerical example below rules out this situation.

[21]That is, unlike the two-stores example previously, we did not start by computing a market equilibrium for given parameter values. Rather we chose the prices and market shares in Table 4 arbitrarily, and use our approach to determine whether the utility parameters are point- or partially-identified.

versions of the MSA algorithm, to determine whether the identified set of utilities is unitary or a nondegenerate set. In Table 5, we report the results.

Table 4. Mixture vertical differentiation, 5 stores and 8 goods

Brand	p_1	p_2	p_3	p_4	p_5	MktShare
A	3.32	3.36	3.45	3.37	3.35	0.07
B	3.88	3.60	3.53	3.39	3.07	0.06
C	3.70	3.30	4.16	4.31	4.25	0.20
D	3.98	4.12	4.06	3.11	4.09	0.39
E	4.20	4.34	4.21	4.29	4.35	0.16
F	4.49	4.82	4.25	3.73	4.86	0.08
G	7.13	7.92	7.95	7.99	7.71	0.01
H	8.34	8.37	8.59	8.62	8.67	0.05

There are two features to note. First, for all entries, the LB and UB practically coincide, suggesting that the utilities are point identified. Second, recall from our earlier discussion that in NARUM models, the normalization entails some loss of generality, and for that reason we examine how robust our results are to choice of normalization. Table 5 also reports the results for four different choices of the normalizing brand. We see that while the results are numerically different, the utility ranking among the 8 goods – from lowest to highest, this is A,B,C,D,E,F,G,H – is invariant to the normalization.

Table 5. Lower bound and upper bound in the mixture vertical differentiation with 5 stores and 8 goods

Brand	LB	UB	LB	UB	LB	UB	LB	UB
A	0.00	0.00	-0.46	-0.43	-3.27	-3.18	-2.90	-2.88
B	0.43	0.46	0.00	0.00	-2.81	-2.76	-2.44	-2.44
C	2.04	2.06	1.60	1.60	-1.21	-1.21	-0.84	-0.84
D	2.88	2.90	2.44	2.44	-0.37	-0.37	0.00	0.00
E	3.25	3.27	2.82	2.82	0.00	0.00	0.37	0.37
F	3.31	3.33	2.88	2.88	0.06	0.06	0.43	0.43
G	6.49	6.51	6.06	6.06	3.24	3.24	3.62	3.61
H	7.75	7.77	7.31	7.32	4.50	4.50	4.87	4.87

6. Empirical Application: Voting in European Parliament Elections

We illustrate the matching-based algorithm by estimating an aggregate spatial voting model using data from the 1999 Parliamentary Elections in the European Union countries, following Merlo and de Paula (2017).

6.1. **Model.** We consider a spatial voting framework in which both voters and political parties are characterized by their "location" in the political spectrum, which is the Cartesian plane $\mathbb{R}^2$. Voter i is characterized by her ideal point $t_i \in \mathbb{R}^2$ within this space; likewise, whereas candidate (political party) j has an ideological position $C_j \equiv (C_{j1}, C_{j2}) \in \mathbb{R}^2$.

Voters are ideological: voter i from electoral precinct m votes for the party with the platform closest to her ideal point t_{mi}. Specifically, her preferred party is:

$$D_{ri} = \text{argmin}_{j \in J_r} d(t_{mi}, C_{mj}), \tag{6.1}$$

$d(,)$: $\mathbb{R}^2 \to \mathbb{R}_+$ is the distance function, which is assumed quadratic:

$$d(t, C) = (t - C)^{'} W (t - C). \tag{6.2}$$

W is a 2×2 weighting matrix, which for simplicity we assume to be identity: $W = \mathbb{I}_2$. Unlike Merlo and de Paula (2017) who consider the nonparametric identification and estimation of the distribution of voter ideal points t, we assume that they come from a bivariate normal distribution, as follows:

$$\begin{aligned} t_{mi1} &= \left(X_m' \alpha + \epsilon_{mi1}\right), \\ t_{mi2} &= \left(X_m' \beta + \epsilon_{mi2}\right), \end{aligned} \tag{6.3}$$

$$(\epsilon_{mi1}, \epsilon_{mi2}) \sim N(\mathbf{0}, \mathbb{I}_2), \text{ i.i.d. over } i,\ j,\ m \tag{6.4}$$

where $X_m = (x_{mk},\ k = 1, \ldots, K)$, are aggregate demographic and economic variables in precinct m, and α and β denote $K \times 1$ parameter vectors of interest, which we refer to collectively as $\theta = (\alpha, \beta)$. As data we use precinct-level vote shares from the 1999 European Parliament elections. Altogether we have voting data from 822 electoral precincts in 22 regions, which are typically countries but may be sub-national regions distinguished by different sets of political parties.[22] Parties' ideological positions are taken from Hix, Noury, and Roland (2006), who computed two ideological positions for each party: $C = (C_{j1}, C_{j2})$,

[22] The 22 regions are: Austria, Finland, France, Germany, Greece, Italy-Center, Italy-Islands, Italy-Northeast, Italy-Northwest, Italy-South, Portugal, Spain, Sweden, Netherlands, UK-East Midlands, UK-Eastern, UK-London, UK-Northwest, UK-Southeast, UK-Southwest, UK-West Midlands, and UK-Yorkshire.

with C_{j1} denoting position on a left-right spectrum, and C_{j2} denoting party's stance on the EU (with larger values denoting, resp. a more right-wing position and more pro-EU stance). We use $K = 3$ precinct-specific socio-economic and demographic variables: the female-to-male ratio, the proportion of the population older than 35 years, and the unemployment rate.[23] All the data we use are available from the *Review of Economic Studies* website.[24]

For computation, we used Knitro for the general estimation problem, and Gurobi 8.1 for computing the linear programming problem in the inner loop.[25]

By substituting (6.3) into (6.2), the "disutility" that voter i gets from voting for candidate j becomes:

$$
\begin{aligned}
U_{ij} &= d(t_i, C_j) = (t_{i1} - C_{j1})^2 + (t_{i2} - C_{j2})^2 \\
&= (X_m'\alpha)^2 + C_{j1}^2 - 2\left(C_{j1} * (X_m'\alpha)\right) + (X_m'\beta)^2 + C_{j2}^2 - 2\left(C_{j2} * (X_m'\beta)\right) + \\
&\quad \epsilon_{i1}^2 + 2\left(X_m'\alpha - C_{j1}\right)\epsilon_{i1} + \epsilon_{i2}^2 + 2\left(X_m'\beta - C_{j2}\right)\epsilon_{i2} \\
&\equiv \delta_{mj} + g(\epsilon_{i1}, \epsilon_{i2}, X_m, C_j; \theta). \qquad (6.5)
\end{aligned}
$$

where δ_{mj} (corresponding to the second row in the above display) is the "mean utility" of party J in precinct m, and $g(\cdots)$ (corresponding to the third row in the above display) is a composite random error term, consisting of all terms which vary across voters i. This specification of the utility is a variety of the pure-characteristics model, as pointed out by Merlo and de Paula (2017), and we use our "Matching-NFXP" algorithm, which we also used for the simulations in Table 1, for estimation. Essentially, this resembles the Berry, Levinsohn, and Pakes (1995) estimation algorithm except that the inner loop utilizes a matching-based algorithm in place of the contraction mapping in BLP. We describe the outer and inner loops in turn.

6.1.1. *Estimation: Inner loop.* In the inner loop, we will ignore the function form of the mean utilities δ_{mj} as given in Eq. (6.5). Hence, for given parameters θ, will solve a two-sided matching problem in order to recover the qualities δ_{mj} for the parties j competing in each precinct m. As before, for each electoral precinct m, we consider a large but finite

[23]Merlo and de Paula (2017) include additionally GDP as a regressor, and hence exclude Austria and Italy from their analysis due to missing values of this variable.

[24]`https://doi.org/10.1093/restud/rdw046`

[25]After some experiments, we find that Knitro significantly outperformed (in terms of the speed and solution quality) other derivative-free algorithms such as Nelder-Mead and simulated annealing.

matching market involving I voters. (For the results here, we used $I = 500$ voters in each precinct.) For each voter i, we draw (e_{mi1}, e_{mi2}) i.i.d. from a bivariate standard normal distribution, and construct the composite error terms $g(\epsilon_{i1}, \epsilon_{i2}, C_j; \theta)$ using Eq. (6.5).

Following the earlier discussion (cf. Eq. (4.3)), in each call to the inner loop we solve the demand inversion problems across all precincts simultaneously, by combining all the linear programs across all precincts into the following single large linear program:

$$\inf_{u_{mi}, \delta_{mj}, m \in \mathcal{M}} \quad \sum_{m=1}^{M} \left[\sum_{i=1}^{I} \frac{1}{I} u_{mi} - \sum_{j \in \mathcal{J}_m} s_{mj} \delta_{mj} \right] \tag{6.6}$$

$$s.t. \quad u_{mi} - \delta_{mj} \geq -g(\epsilon_{i1}, \epsilon_{i2}, C_j; \theta) \quad \forall m \in \mathcal{M} \tag{6.7}$$

$$\delta_{m1} = 0 \quad \forall m \in \mathcal{M}. \tag{6.8}$$

Let $\left\{ \hat{\delta}_{mj}(\theta) \right\}_{m,j}$ denote the optimized values of the δ_{mj}'s from this problem. (The u_{mi}'s denote the optimized utilities for each of the I simulated voters in precinct m, but are not used in estimation.)

6.1.2. *Estimation: Outer loop.* For the outer loop, we minimize the least-square differences between the $\hat{\delta}_{mj}(\theta)$'s emerging from the inner loop and the functional form for δ_{mj}, as implied in Eq. (6.5). to estimate θ:

$$\min_{\theta} \sum_{m,j} \left[\hat{\delta}_{mj}(\theta) - (X'_m \alpha)^2 + C_{j1}^2 - 2C_{j1}(X'_m \alpha) + (X'_m \beta)^2 + C_{j2}^2 - 2C_{j2}(X'_m \beta) \right]^2 \tag{6.9}$$

6.2. **Results.** Table 6 contains our estimation results.[26] The estimated coefficients summarize the contribution of demographic variables on the two dimensions of voters' ideal points. For the first dimension, the coefficient on unemployment rate is negative (-2.6159), indicating that left-leaning precincts tend to have higher unemployment rates. Precincts with larger female-to-male ratio tend to be more conservative (0.8141), while there is no significant relation between age (measured by the proportion of population older than 35 years) on the tendency to be pro-conservative.

For the second dimension, we find that precincts with higher unemployment rate are strongly less supportive of the EU. Precincts with higher ratio of female voters are significantly more pro-EU under both models—a finding that is consistent with both Merlo and de Paula (2017) and the Eurobarometer surveys. Finally, the share of the population older

[26]Since this is an ARUM, without loss of generality, we normalized $\delta_{m1} = 0$, corresponding to the party listed first in alphabetical order in each precinct m.

TABLE 6. Estimation Results: 1999 European Parliamentary Elections

dim. 1: pro-conservative	
unemployment rate	-2.6159
	(-3.02,-2.17)
female-to-male ratio	0.8141
	(0.5,1.12)
above 35-year-old(%)	-0.0724
	(-0.56,0.49)
dim. 2: pro-EU	
unemployment rate	-2.9525
	(-3.53,-2.28)
female-to-male ratio	0.6187
	(0.29,0.96)
above 35-year-old(%)	-0.0232
	(-0.6,0.55)
observations	822
Checking partial identification:	
$\max_{m,j}\{\delta_{mj}^{UB} - \delta_{mj}^{LB}\}$	1.11×10^{-9}

Note: The inner loop utilizes the linear programming algorithm (Section 4.2.1). We report the 95% bootstrap confidence interval constructed from 500 bootstrap samples. δ_{mj}^{UB} and δ_{mj}^{LB} refer, respectively, to estimates of the upper and lower bound of the identified utility parameters, computed using the algorithm in Eq. (4.2).

than 35 years has a negative impact on pro-EU, but the parameter is insignificant. Altogether, economic considerations – as exemplified in the unemployment rate – appear to be the strongest and most consistent explanator of voters' preferences across European regions; of the two included demographic variables, gender appears to play a more significant role in political attitudes than age.

We also use our algorithms to check whether the demand map is invertible by solving for the upper and lower bounds on the precinct/party qualities δ_{mj} using the linear program in

Eq. (4.2). At the bottom of Table 6, we report the maximal (across all precincts and parties) difference between the estimated upper and lower bounds in the δ_{mj}'s, evaluated at the parameter values reported in Table 6. As is evident, the difference is minuscule, suggesting that partial identification of these1 parameters is not an issue for this model.[27] Finally, the combined linear program in Eq. (6.6) is a big time-saver, as executing the demand inversion problem *simultaneously* across all precincts is ten times faster than performing the demand inversion separately for each precinct. This suggests that simultaneous solution of demand inversion problems is a very practical advantage of the linear programming approach.

7. Additional theoretical results

Here we derive additional theoretical results for the matching model in Section 3.1.

7.1. **Existence.** As our main theoretical results explore a new equivalence between the identified utility set $\tilde{\sigma}^{-1}$ and the equilibrium payoffs in a two-sided matching game, we start by considering the existence of the equilibrium payoff set in this game under our assumptions.[28] In order to show that $\tilde{\sigma}^{-1}(s)$ is non-empty, we need to make slightly stronger assumptions than the ones that were previously imposed. In particular, Assumption 1 will be replaced by the following one:

Assumption 3 (Stronger regularity of $\mathcal{U}$)**.** *Assume:*

(a) for every $\varepsilon \in \Omega$, the map $\varepsilon \mapsto \mathcal{U}_{\varepsilon j}(\delta_j)$ is integrable, and

(b) the random map $\delta_j \mapsto \mathcal{U}_{\varepsilon j}(\delta_j)$ is stochastically equicontinuous.

We also need to keep track of the behavior of $\mathcal{U}_{\varepsilon j}(\delta)$ when δ tends to $-\infty$ or $+\infty$, and for this, we introduce the following assumption:

Assumption 4 (Left and right behavior)**.** *Assume that:*

(a) There is $a > 0$ such that $\mathcal{U}_{\varepsilon j}(\delta)$ converges in probability as $\delta \to -\infty$ towards a random variable dominated by $-a$, that is: for all $\eta > 0$, there is $\delta^* \in \mathbb{R}$ such that $\Pr(\mathcal{U}_{\varepsilon j}(\delta^*) > -a) < \eta$.

[27]When the δ's are partially-identified, the identification and estimation of the structural parameters (the β's and the parameters in the distribution of random coefficients) is an open question, and we do not consider it here. Identification of these parameters typically relies upon instruments, and the associated moment conditions. The literature on identification and inference in moment condition models with possibly partially-identified parameters is still nascent; see Chen, Christensen, and Tamer (2018) for one recent paper.

[28]We note that this result is a contribution to matching theory per se, as it implies the existence of a solution to the equilibrium transport problem, as introduced in Galichon (2016), Definition 10.1.

(b) $\mathcal{U}_{\varepsilon j}(\delta)$ *converges in probability as* $\delta \to +\infty$ *towards* $+\infty$*, that is: for all* $\eta > 0$ *and* $b \in \mathbb{R}$*, there is* $\delta^* \in \mathbb{R}$ *such that* $\Pr(\mathcal{U}_{\varepsilon j}(\delta^*) < b) < \eta$*.*

We define $\mathcal{S}_0^{int} = \{s \in \mathcal{S}_0 : s_j > 0,\ \forall j \in \mathcal{J}_0\}$. We can now prove the existence theorem.

Theorem 3. *Under Assumptions 1, 2, 3, and 4,* $\tilde{\sigma}^{-1}(s)$ *is non-empty for all* $s \in \mathcal{S}_0^{int}$*.*

7.2. **Uniqueness and convergence.** Next we consider uniqueness. Assume the random maps $\delta \mapsto \mathcal{U}_{\varepsilon j}(\delta)$ are invertible for each $\varepsilon \in \Omega$ and $j \in \mathcal{J}$, and define Z to be the random vector such that $Z_j = \mathcal{U}_{\varepsilon j}^{-1}(\mathcal{U}_{\varepsilon 0}(\delta_0))$. Z is a random vector valued in $\mathbb{R}^{\mathcal{J}}$; let P_Z be the probability distribution of Z. We will consider the following assumption on P_Z.

Assumption 5. *Assume that:*

(i) the map $\delta_j \mapsto \mathcal{U}_{\varepsilon j}(\delta_j)$ *is invertible for each* $\varepsilon \in \Omega$ *and* $j \in \mathcal{J}$*, and*

(ii) P_Z *has a non-vanishing density over* $\mathbb{R}^J$*.*

Theorem 4. *Under Assumptions 1, 2 and 5,* $\tilde{\sigma}^{-1}(s)$ *has a single element for all* $s \in \mathcal{S}_0^{int}$*.*

Assumption 5 is actually quite natural. In the case of additive random utility models, the map $\delta_j \mapsto \mathcal{U}_{\varepsilon j}(\delta_j) = \delta_j + \varepsilon_j$ is indeed continuous, and $Z_j = \delta_0 + \varepsilon_0 - \varepsilon_j$ has a non-vanishing density over $\mathbb{R}^J$ when $(\varepsilon_0 - \varepsilon_j)$ does. On the other hand, for Example 2 in Section 5.2, Assumption 5(ii) is violated; because the consumer heterogeneity $\theta \in [0, 1]$, the random vector Z can only have bounded support.[29] Accordingly, this example can have a non-unique identified utility set, as we saw earlier.

Theorem 4 is related to BGH's result on the invertibility of demand systems. Our result is more specialized, applying to discrete-choice demand models with possibly non-additive random errors, and our conditions appear non-nested with those in BGH. In addition, our proof (in the Appendix) is based on showing that the lattice upper and lower bounds coincide under the assumed conditions, which is specific to our matching approach.

Given uniqueness, we also consider convergence properties. While we do not focus on statistical inference in this paper, the next result may be useful for showing asymptotic properties of our procedures. In practice, the vector of market shares s may contain sample uncertainty, and we may approximate P by discretization. This will provide us with a sequence (P^n, s^n) which converges weakly toward (P, s), where P is the true distribution of ε, and s is the vector of market shares in the population. Under assumptions slightly

[29]Specifically, each component Z_j has only bounded support on the real line, $Z_j > p_j - p_0$.

weaker than for Theorem 4, we establish that if P^n and s^n converge weakly to P and s, respectively, then any $\delta^n \in \tilde{\sigma}^{-1}(P^n, s^n)$ will also converge.

Assumption 6. *Assume that:*

(i) the map $\delta_j \mapsto \mathcal{U}_{\varepsilon j}(\delta_j)$ *is invertible for each* $\varepsilon \in \Omega$ *and* $j \in \mathcal{J}$*, and*

(ii) for each $\delta \in \mathbb{R}^{\mathcal{J}}$*, the random vector* $(U_{\varepsilon j}(\delta_j))_{j \in \mathcal{J}}$ *where* $\varepsilon \sim P$ *has a non-vanishing continuous density* $g(u; \delta)$ *such that* $g : \mathbb{R}^{\mathcal{J}} \times \mathbb{R}^{\mathcal{J}} \to \mathbb{R}$ *is continuous.*

Note that Assumption 6 is stronger than Assumption 5.

Theorem 5. *Under Assumptions 1, 2 and 6, assume that* P^n *and* s^n *converge weakly to* P *and* s*, respectively. By theorem 4,* $\tilde{\sigma}^{-1}(P, s)$ *is a singleton, denoted* $\{\delta\}$*. If* $\delta^n \in \tilde{\sigma}^{-1}(P^n, s^n)$ *for all* n*, then* $\delta^n \to \delta$ *holds almost surely.*

8. Related literature

In this section we provide an in-depth discussion of how our paper relates to the existing literature on (1) identification of demand maps and (2) two-sided matching. Theoretically, there is a large literature that tackles the problem of the invertibility of the demand map utilizing either *differentiability* or *monotonicity* of the demand map. The first approach, based on global univalence theorems, requires the *differentiability* of the mapping $\sigma(\delta)$. The Hadamard-Palais univalence theorem (Palais (1959)) asserts that (i) if $\sigma : \mathbb{R}^J \to \mathbb{R}^J$ is C^1, (ii) if its Jacobian is invertible at all points, and (iii) if $\|\sigma(\delta)\| \to \infty$ as $\|\delta\| \to \infty$, then σ is globally invertible; further results are collected in Parthasarathy (1983) and Radulescu and Radulescu (1980). Hadamard-type results therefore permits the study of both existence and uniqueness of a solution to $\sigma(\delta) = s$, and were used by Chiappori and Komunjer (2009) and Kristensen, Nesheim, and de Paula (2014) to study (resp.) multinomial choice and nonadditive random utility models. The results in Gale and Nikaido (1965) focus on uniqueness, leaving existence aside: assuming that the Jacobian of σ is a P-matrix and that the domain is a rectangle, Gale-Nikaido's result guarantees the injectivity of σ – namely, that $\sigma^{-1}(\{s\})$ should have at most one point– but can be empty.

A second approach to demand inversion relies on *gross substitutes*—a form of monotonicity—that σ_j is decreasing, or at least nonincreasing, with respect to $\delta_{j'}$. This category of papers includes the literature on ARUMs, where this property holds automatically (eg. McFadden et al. (1978)). Hotz and Miller (1993) study the problem of the nonemptiness of $\sigma^{-1}(s)$, within general ARUMs; Berry (1994) provided a complete argument (which extends to nonadditive random utility models), and also shows uniqueness of the vector of systematic utilities under continuity conditions. Magnac and Thesmar (2002) investigate identification of structural parameters (period utility flows) in dynamic discrete choice models. Norets and Takahashi (2013) focus on the surjectivity of ARUMs, under the assumption of absolute continuity of the distribution of the additive utility shocks. More broadly, Berry, Gandhi, and Haile (2013) show injectivity in general demand systems with gross substitutes (thus going beyond random utility models) under a connected strong substitute assumption.

Finally we provide a brief and incomplete review of the two-sided matching literature, as one key result in this paper is to show the equivalence between demand inversion and a two-sided matching model. The matching literature is split between models with non-transferable utility (NTU), and transferable utility (TU), with intermediate cases called imperfectly transferable utility (ITU).[30] A connection was made earlier between TU matching models and ARUMs (see Galichon and Salanié (2015) and Chiong, Galichon, and Shum (2016)), and in the present paper, we are making a novel connection between matching models with ITU models and NARUMs.

Broadly speaking, there are 2 classes of algorithms for computing matching models. The first class are "deferred acceptance algorithms" which revolve around Tarski's fixed point theorem, and interpreting stable matchings as fixed points of monotone mappings. They apply to NTU and ITU models, but not to TU models. These ideas appeared in Adachi (2000), followed by Fleiner (2003), Echenique and Oviedo (2004) and Hatfield and Milgrom (2005). The seminal deferred acceptance algorithms (Gale and Shapley (1962); Crawford and Knoer (1981); Kelso and Crawford (1982)) can be interpreted in this way.

The second class of methods are "descent methods," which rely on a reformulation of the problem as a convex optimization problem, and apply only in the TU case. Some of the methods involve coordinate descent (as the auction algorithm of Bertsekas and Castanon (1989)), while some involve gradient descent (as the semi-discrete approach of Aurenhammer (1987)). The linear programming solution to the optimal assignment problem described in Shapley and Shubik (1971) also belongs in this category. The study of rates of convergence of these algorithms have been the subject of intense study; see the book Peyré and Cuturi (2017) and Burkard, Dell'Amico, and Martello (2009) for results and further references.

[30]A key reference for the matching literature (TU, NTU and ITU alike) is Roth and Sotomayor (1992), while Galichon (2016) focuses on the TU case, or equivalently, optimal transport methods.

9. Concluding remarks

In this paper we have explored the intimate connection between discrete choice models and two-sided matching models, and used results from the literature on matching under imperfectly transferable utility to derive identification and estimation procedures for discrete choice models based on the non-additive random utility specification. Although the microeconomics literature distinguishes between "one-sided" and "two-sided" demand problems, our results show that this distinction is immaterial for the purpose of estimating discrete-choice models. Given the matching equivalence, it is as appropriate to consider the discrete choice problem of consumers choosing yogurts as one in which "yogurts choose consumers".

The connection between discrete choice and two-sided matching is a rich one, and we are exploring additional implications. For instance, the phenomenon of "multiple discrete choice" (consumers who choose more than one brand, or choose bundles of products on a purchase occasion) is challenging and difficult to model in the discrete choice framework[31] but is quite natural in the matching context, where "one-to-many" markets are commonplace – perhaps the most prominent and well-studied being the National Residents Matching Program for aspiring doctors in the United States (cf. Roth (1984)). We are exploring this connection in ongoing work.

References

Ackerberg, D. A., and M. Rysman (2005): "Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects," *RAND Journal of Economics*, 36(4), 771–788.

Adachi, H. (2000): "On a characterization of stable matchings," *Economics Letters*, 68, 43–49.

Aguirregabiria, V., and P. Mira (2002): "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models," *Econometrica*, 70(4), 1519–1543.

Apesteguia, J., and M. A. Ballester (2014): "Discrete choice estimation of risk aversion," working paper, Pompeu Fabra.

Arcidiacono, P., and R. A. Miller (2011): "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 79(6), 1823–1867.

Aurenhammer, F. (1987): "Power diagrams: properties, algorithms and applications," *SIAM Journal on Computing*, 16(1), 78–96.

Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating dynamic models of imperfect competition," *Econometrica*, 75(5), 1331–1370.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81(5), 2087–2111.

[31]See Hendel (1999), Dubé (2004), Fox and Bajari (2013) for some applications.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, pp. 841–890.

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

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

Berry, S. T., and P. Haile (2015): "Identification in Differentiated Products Markets," Discussion paper, National Bureau of Economic Research.

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

Bertsekas, D. P. (1992): "Auction algorithms for network flow problems: A tutorial introduction," *Computational optimization and applications*, 1(1), 7–66.

Bertsekas, D. P., and D. A. Castanon (1989): "The auction algorithm for the transportation problem," *Annals of Operations Research*, 20(1), 67–96.

Bonnet, O., A. Galichon, Y.-W. Hsieh, K. O'Hara, and M. Shum (2018a): "Auction.Rcpp - R Wrapper to General Auction and Assignment Methods," `https://github.com/TraME-Project/auction`.

Bonnet, O., A. Galichon, Y.-W. Hsieh, K. O'Hara, and M. Shum (2018b): "Yogurts - Fast IPFP and MSA Methods for R," `https://github.com/TraME-Project/yogurts`.

Bresnahan, T. F. (1981): "Departures from marginal-cost pricing in the American automobile industry: Estimates for 1977–1978," *Journal of Econometrics*, 17(2), 201–227.

Burkard, R., M. Dell'Amico, and S. Martello (2009): *Assignment Problems*. SIAM.

Chen, X., T. M. Christensen, and E. Tamer (2018): "Monte Carlo confidence sets for identified sets," *Econometrica*, 86(6), 1965–2018.

Chernozhukov, V., A. Galichon, M. Henry, and B. Pass (2019): "Single market nonparametric identification of multi-attribute hedonic equilibrium models," *arXiv preprint arXiv:1709.09570*.

Chiappori, P., and I. Komunjer (2009): "On the Nonparametric Identification of Multiple Choice Models," working paper, Columbia.

Chiong, K. X., A. Galichon, and M. Shum (2016): "Duality in dynamic discrete choice models," *Quantitative Economics*, 7, 83–115.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Muiltiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

Cohen, A., and L. Einav (2007): "Estimating risk preferences from deductible choice," *American economic review*, 97(3), 745–788.

Crawford, V. P., and E. M. Knoer (1981): "Job matching with heterogeneous firms and workers," *Econometrica*, pp. 437–450.

Demange, G., and D. Gale (1985): "The strategy structure of two-sided matching markets," *Econometrica*, pp. 873–888.

Dempe, S., V. Kalasshnikov, G. A. Pérez-Valdés, and N. Kalashnykova (2015): *Bilevel Programming Problems: Theory, Algorithms and Applications to Energy Networks*. Springer-Verlag.

Dubé, J.-P. (2004): "Multiple discreteness and product differentiation: Demand for carbonated soft drinks," *Marketing Science*, 23(1), 66–81.

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.

Echenique, F., and J. Oviedo (2004): "Core Many-to-one Matchings by Fixed Point Methods," *Journal of Economic Theory*, 115, 358–376.

Fleiner, T. (2003): "A fixed-point approach to stable matchings and some applications," *Mathematics of Operations Research*, 28.1, 103–126.

Fox, J. T., and P. Bajari (2013): "Measuring the efficiency of an FCC spectrum auction," *American Economic Journal: Microeconomics*, 5(1), 100–146.

Gale, D., and H. Nikaido (1965): "The Jacobian matrix and global univalence of mappings.," *Mathematical Annals*, 159, 81–93.

Gale, D., and L. S. Shapley (1962): "College admissions and the stability of marriage," *The American Mathematical Monthly*, 69(1), 9–15.

Galichon, A. (2016): *Optimal Transport Methods in Economics*. Princeton.

Galichon, A., and Y.-W. Hsieh (2019): "Aggregate Stable Matching with Money Burning," *working paper*.

Galichon, A., and B. Salanié (2015): "Cupid's invisible hand: Social surplus and identification in matching models," working paper, NYU.

Hatfield, J. W., and P. R. Milgrom (2005): "Matching with contracts," *American Economic Review*, pp. 913–935.

Hendel, I. (1999): "Estimating multiple-discrete choice models: An application to computerization returns," *The Review of Economic Studies*, 66(2), 423–446.

Hix, S., A. Noury, and G. Roland (2006): "Dimensions of Politics in the European Parliament," *American Journal of Political Science*, 50, 494–511.

Hotz, V. J., and R. A. Miller (1993): "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, 60(3), 497–529.

Jia-Barwick, P. (2008): "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 76(6), 1263–1316.

Kelso, A. S., and V. P. Crawford (1982): "Job matching, coalition formation, and gross substitutes," *Econometrica*, pp. 1483–1504.

Khan, S., F. Ouyang, and E. Tamer (2016): "Adaptive Rank Inference in Semiparametric Multinomial Response Models," working paper, Boston College.

Kitagawa, J., Q. Mérigot, and B. Thibert (2016): "Convergence of a Newton algorithm for semi-discrete optimal transport," *Forthcoming, Journal of the European Math Society (JEMS)*.

Kristensen, D., L. Nesheim, and A. de Paula (2014): "CCP and the Estimation of Nonseparable Dynamic Models," Working paper, UCL.

Lévy, B. (2015): "A numerical algorithm for L2 semi-discrete optimal transport in 3D," *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6), 1693–1715.

——— (2018): "Geogram - A Programming Library of Geometric Algorithms," .

Lévy, B., and K. O'Hara (2018): "Rgeogram - R Wrapper to the Geogram Library," .

Magnac, T., and D. Thesmar (2002): "Identifying dynamic discrete decision processes," *Econometrica*, 70(2), 801–816.

Mangasarian, O. (1969): *Nonlinear Programming*. Society for Industrial and Applied Mathematics.

Matzkin, R. (2007): "Nonparametric Identification," *Handbook of Econometrics*, 6b.

McFadden, D., et al. (1978): *Modelling the choice of residential location*. Institute of Transportation Studies, University of California California.

McFadden, D., and K. Train (2000): "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15(5), 447–470.

Merlo, A., and A. de Paula (2017): "Identification and Estimation of Preference Distributions When Voters Are Ideological," *Review of Economic Studies*, 84, 1238–1263.

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

Norets, A., and S. Takahashi (2013): "On the surjectivity of the mapping between utilities and choice probabilities," *Quantitative Economics*, 4(1), 149–155.

Nosko, C. (2010): "Competition and quality choice in the cpu market," working paper, Harvard University.

Pakes, A., and J. Porter (2015): "Moment Inequalities for Multinomial Choice with Fixed Effects," *working paper*.

Palais, R. S. (1959): "Natural operations on differential forms," *Transactions of the American Mathematical Society*, 92(1), 125–141.

Pang, J.-S., C.-L. Su, and Y.-C. Lee (2015): "A constructive approach to estimating pure characteristics demand models with pricing," *Operations Research*, 63(3), 639–659.

Parthasarathy, T. (1983): *On Global Univalence Theorems*. Springer.

Peyré, G., and M. Cuturi (2017): *Computational Optimal Transport*. Forthcoming.

Prescott, E. C., and M. Visscher (1977): "Sequential location among firms with foresight," *The Bell Journal of Economics*, pp. 378–393.

Radulescu, M., and S. Radulescu (1980): "Global inversion theorems and applications to differential equations.," *Nonlinear Analysis: Theory, Methods & Applications*, 92, 951–965.

Roth, A. E. (1984): "The evolution of the labor market for medical interns and residents: a case study in game theory," *The Journal of Political Economy*, pp. 991–1016.

Roth, A. E., and M. A. O. Sotomayor (1992): *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press.

Shapley, L. S., and M. Shubik (1971): "The assignment game I: The core," *International Journal of Game Theory*, 1(1), 111–130.

Song, M. (2007): "Measuring Consumer Welfare in the CPU Market: an Application of the Pure-Characteristics Demand Model," *RAND Journal of Economics*, 38(2), 429–446.

Strassen, V. (1965): "The existence of probability measures with given marginals," *Annals of Mathematical Statistics*, 36, 423–439.

Topkis, D. (1998): *Supermodularity and Complementarity*. Princeton University Press.

Train, K. E. (2009): *Discrete Choice Methods with Simulation*. Cambridge university press.

Veinott, A. (2005): "Lecture notes in Supply-Chain optimization," Stanford University.

Vohra, R. V. (2004): *Advanced mathematical economics*. Routledge.

Walsh, J., and L. Dieci (2017): "General auction method for real-valued optimal transport," *arXiv preprint arXiv:1705.06379*.

(Intended for online publication)

Appendix A. Proofs and additional results

We shall prove stronger versions of theorems 1–4 in a more general setting where we allow for possible ties, that is, when Assumption 2 is dropped. In this case, the demand map σ is no longer defined, but should be replaced by a set-valued analog.[32] We first introduce the relevant object, the demand correspondence, before stating our equivalence theorem in this more general setting, and the properties of the inverse demand correspondence.

A.1. **The demand correspondence.** When we drop Assumption 2, indifference between two alternatives may occur with positive probability, and the random set of alternatives preferred by the agent may contain several elements. This is for instance the case in an additive random utility model with discrete shocks (see below). Hence one cannot define a map σ by (2.2). Instead, we can define the *demand correspondence* $\Sigma(\delta)$ at vector δ as the set of market shares compatible with the optimal choices of consumers when the systematic utilities are δ and some tie-breaking rule is arbitrarily chosen. That is:

Definition 4 (Demand correspondence)**.** The demand correspondence $\Sigma : \mathbb{R}^{\mathcal{J}_0} \to \mathcal{P}(\mathcal{S}_0)$ is a function from $\mathbb{R}^{\mathcal{J}_0}$ to the power set of $\mathcal{S}_0$ such that $\Sigma(\delta)$, is the set of market shares s such that there is a random variable $\tilde{j}$ valued in $\mathcal{J}_0$ with probability mass vector s, and such that $\tilde{j}$ maximizes $U_{\varepsilon j}(\delta_j)$ over $j \in \mathcal{J}_0$ almost surely.

We define the inverse demand correspondence by

$$\Sigma^{-1}(s) = \left\{\delta \in \mathbb{R}^{\mathcal{J}_0} : s \in \Sigma(\delta)\right\}$$

which is the set of utility vectors δ that rationalize the vector of market shares $s \in \mathbb{R}^{\mathcal{J}_0}$, which is the identified set of utilities.

Example A.1. Consider the case of an ARUM without heterogeneity: $\varepsilon_j = 0$ a.s. for $j \in \mathcal{J}_0$, so that $\mathcal{U}_{\varepsilon_i j}(\delta_j) = \delta_j$. Then $\Sigma(\delta)$ contains a single element or multiple elements depending on the number of elements contained in $\arg\max_{j \in \mathcal{J}_0} \delta_j$. If the argmax has a single element j^*, then $\Sigma(\delta)$ is the $s \in \mathcal{S}_0$ such that $s_{j^*} = 1$ and $s_j = 0$ for $j \neq j^*$. If the argmax is a set B with multiple elements, then $\Sigma(\delta)$ is the set of $s \in \mathbb{R}_+^{\mathcal{J}_0}$ such that $\sum_{j \in J_0} s_j = 1$, and $j \notin B$ implies $s_j = 0$.

The following result is a direct consequence of Strassen's theorem (Strassen (1965), theorem 5). It provides a convenient reexpression of $\Sigma(\delta)$.

[32]See also Pakes and Porter (2015) for a discussion of identification and inference in ARUM models with possible indifferences.

Proposition 1. *Let* $s \in \mathbb{R}_+^{\mathcal{J}_0}$ *be such that* $\sum_{j\in\mathcal{J}_0} s_j = 1$. *Then under Assumption 1, the following statements are equivalent:*

(i) $s \in \Sigma(\delta)$, *and*

(ii) $\forall B \subseteq \mathcal{J}_0,\ \sum_{j\in B} s_j \leq P\left(\max_{j\in B} U_{\varepsilon j}(\delta_j) \geq \max_{j\in\mathcal{J}_0\setminus B} U_{\varepsilon j}(\delta_j)\right)$.

Proof. Direct implication: Let s be the probability mass vector of a random variable $\tilde{j}$ valued in $\mathcal{J}_0$ such that $\tilde{j} \in J(\varepsilon)$. Then for all $B \subseteq \mathcal{J}_0$, one has

$$\sum_{j\in B} s_j = \Pr\left(\tilde{j} \in B\right) \leq \Pr\left(J(\varepsilon) \cap B \neq \emptyset\right). \tag{A.1}$$

Converse implication: Assume (A.1). Then by Strassen's theorem (Strassen (1965), theorem 5), one can construct $\tilde{j}$ and ε on the same probability space such that $\tilde{j} \in J(\varepsilon)$ almost surely.[33] □

To gain some intuition for (ii), note that the RHS of the inequality is the probability that, for all ε such that the optimizing choices contain some alternative(s) in a set B, those alternatives in B are chosen. This is a upper bound on the actual markets for alternatives in set B.[34]

Remark A.1. A necessary condition for the second statement of proposition 1 to hold is

$$s_j \leq P\left(U_{\varepsilon j}(\delta_j) \geq \max_{j'\in\mathcal{J}_0} U_{\varepsilon j'}(\delta_{j'})\right), \quad \text{for all } j \in \mathcal{J}_0 \tag{A.2}$$

which amounts to checking part (ii) in Proposition 1 on the class of singleton subsets. However, this condition is *not* sufficient as shown in the following example. Consider the case when $\mathcal{J}_0$ has three elements and the set $J(\varepsilon)$ of optimal alternatives is $\{j_1\}$ wp 1/3, $\{j_1, j_2\}$ wp 1/3, and $\{j_3\}$ wp 1/3. Then $s = (2/3, 1/3, 0)$ satisfies inequalities (A.2) for every $j \in \mathcal{J}_0$. However there is no random variable $\tilde{j}$ valued in $\mathcal{J}_0$ such that $\tilde{j} \in J(\varepsilon)$. Indeed, if this were the case, $\Pr\left(\tilde{j} = j_3 | J(\varepsilon) = \{j_3\}\right) = 1$, thus $\Pr\left(\tilde{j} = j_3\right) \geq \Pr\left(J(\varepsilon) = \{j_3\}\right) = 1/3$, a contradiction. □

Remark A.2. In general $\Sigma(\delta) \subseteq \left[\underline{\sigma}_j(\delta), \overline{\sigma}_j(\delta)\right]$, where we define

$$\begin{aligned} \underline{\sigma}_j(\delta) &= P\left(\varepsilon \in \Omega : \mathcal{U}_{\varepsilon j}(\delta_j) > \max_{j'\in\mathcal{J}_0\setminus\{j\}} \mathcal{U}_{\varepsilon j}(\delta_{j'})\right) \\ \overline{\sigma}_j(\delta) &= P\left(\varepsilon \in \Omega : \mathcal{U}_{\varepsilon j}(\delta_j) \geq \max_{j'\in\mathcal{J}_0\setminus\{j\}} \mathcal{U}_{\varepsilon j}(\delta_{j'})\right) \end{aligned}$$

For instance, if $\mathcal{J} = \{j_1, j_2\}$ and $\mathcal{U}_{\varepsilon j}(\delta_{j_1}) = \mathcal{U}_{\varepsilon j}(\delta_{j_2}) = \mathcal{U}_{\varepsilon j}(\delta_{j_0})$ for every $\varepsilon \in \Omega$, then $\Sigma(\delta) = \left\{(s_1, s_2) \in \mathbb{R}_+^2 : s_1 + s_2 \leq 1\right\}$; indeed, in this case, the agent is indifferent between the three alternatives in every state of the world, thus any randomized choice is a solution. In this case $\underline{\sigma}_j(\delta) = 0$ and $\overline{\sigma}_j(\delta) = 1$. □

[33] Strassen's theorem is essentially a continuous extension of Hall's marriage theorem.

[34] A similar inequality is used to generate the upper bound choice probabilities in Ciliberto and Tamer (2009)'s study of multiple equilibria in airline entry games.

Remark A.3. Under Assumptions 1 and 2, it follows from proposition 1 that Σ is point-valued, that is $\Sigma(\delta) = \sigma(\delta)$ for all δ. However, it does not mean that $\sigma^{-1}(s)$ is itself point valued. □

Normalization. Just as in section 3, normalization issues will play an important role in our analysis. As a result, we introduce $\tilde{\Sigma}$ as the correspondence $\mathbb{R}^{\mathcal{J}} \to \mathcal{P}(\mathcal{S})$ induced by Σ under normalization $\delta_0 = 0$, hence

$$\tilde{\Sigma}^{-1}(s) = \left\{\delta \in \mathbb{R}^{\mathcal{J}} : s \in \Sigma(\delta') \text{ for } \delta' \in \mathbb{R}^{\mathcal{J}_0} \text{ with } \delta'_{-0} = \delta \text{ and } \delta'_0 = 0\right\}.$$

A.2. **Proof of theorems 1–4.**

A.2.1. *Proof of theorem 1.* The proof of theorem 1 follows from the following stronger result, where we have removed Assumption 2. We state and proof this stronger result.

Theorem 1'. *Under Assumption 1, consider a vector of market shares s that satisfies $s_j > 0$ and $\sum_{j\in\mathcal{J}_0} s_j = 1$. Consider a vector $\delta \in \mathbb{R}^{\mathcal{J}_0}$. Then, the two following statements are equivalent:*

(i) δ belongs to the identified utility set $\Sigma^{-1}(s) = \left\{\delta \in \mathbb{R}^{\mathcal{J}_0} : s \in \Sigma(\delta)\right\}$ associated with s in the sense of Definition 2 in the discrete choice problem with $\varepsilon \sim P$, and

(ii) there exists $\pi \in \mathcal{M}(P, s)$ and $u = \max_{j\in\mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j)$ defined by (2.1) such that $(\pi, u, -\delta)$ is an equilibrium outcome in the sense of Definition 3 in the matching problem, where

$$f_{\varepsilon j}(u) = u \text{ and } g_{\varepsilon j}(-\delta) = -\mathcal{U}_{\varepsilon j}(\delta). \tag{A.3}$$

Proof (a) From demand inversion to equilibrium matching: Consider $\delta \in \tilde{\Sigma}^{-1}(s)$ a solution to the demand inversion problem. Then $s_j = P(\varepsilon \in \Omega : \mathcal{U}_{\varepsilon j}(\delta_j) \geq u(\varepsilon))$, where

$$u(\varepsilon) = \max_{j\in\mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j).$$

Let us show that we can construct π and set $v = -\delta$ such that (π, u, v) is an equilibrium outcome, which is to say it satisfies the three conditions of Definition 3.

Let us introduce $J(\varepsilon) = \arg\max_{j\in\mathcal{J}_0} \{\mathcal{U}_{\varepsilon j}(\delta_j)\}$ the set of yogurts that maximize consumer ε's utility. Then $\sigma_j(\delta) = \Pr(j \in J(\varepsilon))$. Let us show that $J(\varepsilon)$ has one element with probability one. Indeed, otherwise, $\{j, j'\} \subseteq J(\varepsilon)$ would arise with positive probability for some pair $j \neq j'$, which would imply that there is a positive probability of indifference between j and j', in contradiction with (2.2). Hence for each open set B,

$$\Pr(J(\varepsilon) \cap B \neq \emptyset) = \sum_{j\in B} \Pr(J(\varepsilon) = \{j\}) = \sum_{j\in B} \Pr(j \in J(\varepsilon)) = \sigma_j(\delta).$$

In particular, for each open set B, one has

$$s(B) \leq \Pr(J(\varepsilon) \cap B \neq \emptyset).$$

By Strassen's theorem (Strassen (1965), theorem 5), this implies that there is a probability distribution $\pi \in \mathcal{M}(P,s)$ such that $j \in J(\varepsilon)$ on the support of π. Hence π satisfy condition (i) in Definition 3. But $j \in J(\varepsilon)$ implies $\mathcal{U}_{\varepsilon j}(\delta_j) = u(\varepsilon)$. Introducing $v(j) = -\delta_j$, $g_{\varepsilon j}(v(j)) = -\mathcal{U}_{\varepsilon_i j}(\delta_j)$, and $f_{\varepsilon j}(x) = x$, one has

$$f_{\varepsilon j}(u(\varepsilon)) + g_{\varepsilon j}(v(j)) \geq 0$$

for all (ε, j), with equality on the support of π. Hence, conditions (ii) and (iii) in Definition 3 are met, and (π, u, v) is an equilibrium outcome.

(b) From equilibrium matching to demand inversion: Let (π, u, v) be an equilibrium matching in the sense of Definition 3, where $f_{\varepsilon j}(x) = x$ and $g_{\varepsilon j}(y) = -\mathcal{U}_{\varepsilon_i j}(-y)$. Then letting $\delta = -v$, one has by condition (ii) that for any $\varepsilon \in \Omega$ and $j \in \mathcal{J}_0$,

$$u(\varepsilon) - \mathcal{U}_{\varepsilon j}(\delta_j) \geq 0$$

thus $u(\varepsilon) \geq \max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j)$. But by condition (iii), for $j \in Supp(\pi(.|\varepsilon))$, one has $u(\varepsilon) = \mathcal{U}_{\varepsilon j}(\delta_j)$, thus

$$u(\varepsilon) = \max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j).$$

Condition (iii) implies that if $(\tilde{\varepsilon}, \tilde{j}) \sim \pi$, then $\Pr(J(\tilde{\varepsilon}) = \{\tilde{j}\}) = 1$, thus

$$\sigma_j(\delta) = P(\varepsilon \in \Omega : J(\tilde{\varepsilon}) = \{j\}) = \Pr(\tilde{j} = j) = s_j.$$

Hence $s \in \tilde{\Sigma}(\delta)$, QED.

A.2.2. *Proof of theorem 2.* Again, we state and prove a version of theorem 2 where Assumption 2 has been removed.

Theorem 2' (Inverse isotonicity of demand)**.** *Under Assumption 1, consider s and s' in $\mathcal{S}_0$ such that $s_j \leq s'_j$ for all $j \in \mathcal{J}$. If there are two vectors δ and δ' satisfying (2.3) such that $s \in \tilde{\Sigma}(\delta)$ and $s' \in \tilde{\Sigma}(\delta')$, then*

$$s \in \tilde{\Sigma}(\delta \wedge \delta') \text{ and } s' \in \tilde{\Sigma}(\delta \vee \delta').$$

Remark A.4. To the best of our knowledge, this result is novel in the theory of two-sided matchings with imperfectly transferable utility. While in the case of matching with (perfectly) transferable utility, it follows easily from the fact that the value of the optimal assignment problem is a supermodular function in $(P, -s)$, (see e.g. Vohra (2004), theorem 7.20), to the best of our knowledge it is novel beyond that case[35]. □

Proof. Assume $s_j \leq s'_j$ for all $j \in \mathcal{J}$, and let δ and δ' in satisfying (2.3) such that $s \in \tilde{\Sigma}(\delta)$ and $s' \in \tilde{\Sigma}(\delta')$. Let $u(\varepsilon) = \max_{j \in \mathcal{J}_0} U_{\varepsilon j}(\delta_j)$ and $u'(\varepsilon) = \max_{j \in \mathcal{J}_0} U_{\varepsilon j}(\delta'_j)$. Let $\delta^\wedge = \delta \wedge \delta'$ and $\delta^\vee = \delta \vee \delta'$ i.e.

$$\delta^\wedge_j = \min(\delta_j, \delta'_j) \text{ and } \delta^\vee_j = \max(\delta_j, \delta'_j),$$

[35]Demange and Gale (1985) show isotonicity in the strong set order with respect to reservation utilities, which is a different result.

and let

$$u^{\wedge}(\varepsilon)=\min\left(u(\varepsilon),u'(\varepsilon)\right) \text{ and } u^{\vee}(\varepsilon)=\max\left(u(\varepsilon),u'(\varepsilon)\right).$$

(a) Proof of $s\in\tilde{\Sigma}(\delta\wedge\delta')$: By Strassen's theorem, the fact that $s\in\tilde{\Sigma}(\delta)$ and $s'\in\tilde{\Sigma}(\delta)$ is equivalent to the fact that for all $A\subseteq\mathcal{J}_0$,

$$\sum_{j\in A}s_j \leq P\left\{\varepsilon\in\Omega:\exists j\in A,\ u(\varepsilon)=U_{\varepsilon j}(\delta_j)\right\}, \text{ and} \tag{A.4}$$

$$\sum_{j\in A}s'_j \leq P\left\{\varepsilon\in\Omega:\exists j\in A,\ u'(\varepsilon)=U_{\varepsilon j}\left(\delta'_j\right)\right\}. \tag{A.5}$$

By the converse implication in Strassen's theorem, in order to show that $s\in\tilde{\Sigma}(\delta\wedge\delta')$, it is sufficient to show that for all $A\subseteq\mathcal{J}_0$,

$$\sum_{j\in A}s_j\leq P\left\{\varepsilon\in\Omega:\exists j\in A,\ u^{\wedge}(\varepsilon)=U_{\varepsilon j}\left(\delta_j^{\wedge}\right)\right\}. \tag{A.6}$$

Take $A\subseteq\mathcal{J}_0$, and let

$$A^{>}=\left\{j\in A:\delta_j>\delta'_j\right\} \text{ and } A^{\leq}=\left\{j\in A:\delta_j\leq\delta'_j\right\}$$

while one defines

$$\Omega^{>}=\left\{\varepsilon\in\Omega:u(\varepsilon)>u'(\varepsilon)\right\} \text{ and } \Omega^{\leq}=\left\{\varepsilon\in\Omega:u(\varepsilon)\leq u'(\varepsilon)\right\}.$$

By (A.5) applied to $A=A^{>}$, one has

$$\sum_{j\in A^{>}}s_j\leq\sum_{j\in A^{>}}s'_j\leq P\left\{\varepsilon\in\Omega:\exists j\in A^{>},\ u'(\varepsilon)=U_{\varepsilon j}\left(\delta'_j\right)\right\};$$

but if $j\in A^{>}$ and if $u'(\varepsilon)=U_{\varepsilon j}\left(\delta'_j\right)$, then $\varepsilon\in\Omega^{>}$. Indeed, otherwise one would have $U_{\varepsilon j}(\delta_j)\leq u(\varepsilon)\leq u'(\varepsilon)=U_{\varepsilon j}\left(\delta'_j\right)$, which would contradict $\delta_j>\delta'_j$. Hence, the latter display implies

$$\sum_{j\in A^{>}}s_j\leq P\left\{\varepsilon\in\Omega^{>}:\exists j\in A^{>},\ u'(\varepsilon)=U_{\varepsilon j}\left(\delta'_j\right)\right\}$$

thus

$$\sum_{j\in A^{>}}s_j\leq P\left\{\varepsilon\in\Omega^{>}:\exists j\in A,\ u^{\wedge}(\varepsilon)=U_{\varepsilon j}\left(\delta_j^{\wedge}\right)\right\}. \tag{A.7}$$

By (A.4) applied to $A=A^{\leq}$, one has

$$\sum_{j\in A^{\leq}}s_j\leq P\left\{\varepsilon\in\Omega:\exists j\in A^{\leq},\ u(\varepsilon)=U_{\varepsilon j}(\delta_j)\right\};$$

but if $j \in A^{\leq}$ and if $u(\varepsilon) = U_{\varepsilon j}(\delta_j)$, then $\varepsilon \in \Omega^{\leq}$. Indeed, otherwise one would have $U_{\varepsilon j}(\delta_j) = u(\varepsilon) > u'(\varepsilon) \geq U_{\varepsilon j}\left(\delta'_j\right)$, which would contradict $\delta_j \leq \delta'_j$. Thus the latter display implies

$$\sum_{j \in A^{\leq}} s_j \leq P\left\{\varepsilon \in \Omega^{\leq} : \exists j \in A^{\leq},\ u(\varepsilon) = U_{\varepsilon j}(\delta_j)\right\};$$

hence

$$\sum_{j \in A^{\leq}} s_j \leq P\left\{\varepsilon \in \Omega^{\leq} : \exists j \in A,\ u^{\wedge}(\varepsilon) = U_{\varepsilon j}\left(\delta_j^{\wedge}\right)\right\}. \tag{A.8}$$

By summation of (A.7) and (A.8), one obtains (A.6), and hence

$$s \in \tilde{\Sigma}(\delta \wedge \delta'),\ QED.$$

(b) Proof of $s' \in \tilde{\Sigma}(\delta \vee \delta')$: By Strassen's theorem, the fact that $s \in \tilde{\Sigma}(\delta)$ and $s' \in \tilde{\Sigma}(\delta)$ is equivalent to the fact that for any Borel subset $B \subseteq \Omega$,

$$P(B) \leq \sum_{j \in \mathcal{J}_0} s_j 1\left\{\exists \varepsilon \in B : u(\varepsilon) = U_{\varepsilon j}(\delta_j)\right\}, \text{ and} \tag{A.9}$$

$$P(B) \leq \sum_{j \in \mathcal{J}_0} s'_j 1\left\{\exists \varepsilon \in B : u'(\varepsilon) = U_{\varepsilon j}\left(\delta'_j\right)\right\}. \tag{A.10}$$

By the converse of Strassen's theorem, in order to show that $s' \in \tilde{\Sigma}(\delta \vee \delta')$, it is sufficient to show that for any $B \subseteq \Omega$,

$$P(B) \leq \sum_{j \in \mathcal{J}_0} s'_j 1\left\{\exists \varepsilon \in B : u^{\vee}(\varepsilon) = U_{\varepsilon j}\left(\delta_j^{\vee}\right)\right\}. \tag{A.11}$$

Take a Borel subset $B \subseteq \Omega$, and let

$$B^{>} = \{\varepsilon \in B : u(\varepsilon) > u'(\varepsilon)\} \text{ and } B^{\leq} = \{\varepsilon \in B : u(\varepsilon) \leq u'(\varepsilon)\}$$

while one defines

$$\mathcal{J}_0^{>} = \left\{j \in \mathcal{J}_0 : \delta_j > \delta'_j\right\} \text{ and } \mathcal{J}_0^{\leq} = \left\{j \in \mathcal{J}_0 : \delta_j \leq \delta'_j\right\}.$$

By (A.9) applied to $B = B^{>}$, one has

$$P\left(B^{>}\right) \leq \sum_{j \in \mathcal{J}_0} s_j 1\left\{\exists \varepsilon \in B^{>} : u(\varepsilon) = U_{\varepsilon j}(\delta_j)\right\}; \tag{A.12}$$

but if $\varepsilon \in B^{>}$ and $u(\varepsilon) = U_{\varepsilon j}(\delta_j)$, then $j \in \mathcal{J}_0^{>}$; otherwise $\delta_j \leq \delta'_j$, and thus $u'(\varepsilon) < u(\varepsilon) \leq U_{\varepsilon j}\left(\delta'_j\right)$, a contradiction. Hence, the sum on the right hand-side of (A.12) can be restricted to the elements of $\mathcal{J}_0^{>}$, which implies

$$\begin{aligned} P\left(B^{>}\right) &\leq \sum_{j \in \mathcal{J}_0^{>}} s_j 1\left\{\exists \varepsilon \in B : u(\varepsilon) = U_{\varepsilon j}(\delta_j)\right\} \qquad \text{(A.13)} \\ &= \sum_{j \in \mathcal{J}_0^{>}} s_j 1\left\{\exists \varepsilon \in B : u^{\vee}(\varepsilon) = U_{\varepsilon j}\left(\delta_j^{\vee}\right)\right\}, \end{aligned}$$

thus, using the fact that $s_j \leq s'_j$ for all $j \in \mathcal{J}_0^{>}$, we deduce that

$$P\left(B^{>}\right) \leq \sum_{j \in \mathcal{J}_0^{>}} s'_j 1\left\{\exists \varepsilon \in B : u^{\vee}(\varepsilon) = U_{\varepsilon j}\left(\delta_j^{\vee}\right)\right\}. \tag{A.14}$$

Next, taking $B = B^{\leq}$ in (A.10) implies that

$$P\left(B^{\leq}\right) \leq \sum_{j \in \mathcal{J}_0} s'_j 1\left\{\exists \varepsilon \in B^{\leq} : u'(\varepsilon) = U_{\varepsilon j}\left(\delta'_j\right)\right\}; \tag{A.15}$$

but if $\varepsilon \in B^{\leq}$ and $u'(\varepsilon) = U_{\varepsilon j}\left(\delta'_j\right)$, then $j \in \mathcal{J}_0^{\leq}$; otherwise $\delta_j > \delta'_j$, and thus $U_{\varepsilon j}(\delta_j) > U_{\varepsilon j}\left(\delta'_j\right) = u'(\varepsilon) \geq u(\varepsilon)$, another contradiction. Hence, (A.15) implies

$$\begin{aligned} P\left(B^{\leq}\right) &\leq \sum_{j \in \mathcal{J}_0^{\leq}} s'_j 1\left\{\exists \varepsilon \in B : u'(\varepsilon) = U_{\varepsilon j}\left(\delta'_j\right)\right\} \quad \text{(A.16)} \\ &= \sum_{j \in \mathcal{J}_0^{\leq}} s'_j 1\left\{\exists \varepsilon \in B : u^{\vee}(\varepsilon) = U_{\varepsilon j}\left(\delta_j^{\vee}\right)\right\}. \end{aligned}$$

By summation of (A.14) and (A.16), one obtains (A.11), and thus

$$s' \in \tilde{\Sigma}\left(\delta \vee \delta'\right), \ QED.$$

□

Let us explore what theorem 2 means in the familiar case of ARUMs, which is equivalent to a model of matching with transferable utility.

Example 2.1 Continued. *In the ARUM case (but in that case only), theorem 2 follows from Topkis' (1998) theorem. In this case, Chiong, Galichon, and Shum (2016) have shown that for* $s \in \mathcal{S}_0$

$$\tilde{\Sigma}^{-1}(s) = \arg\max_{\delta : \delta_0 = 0} \left\{ \sum_{j \in \mathcal{J}} \delta_j s_j - \mathbb{E}_P\left[\max_{k \in \mathcal{J}_0} \{\delta_k + \varepsilon_k\}\right] \right\}.$$

Note that $\delta \rightarrow \mathbb{E}_P\left[\max_{k \in J_0}\{\delta_k + \varepsilon_k\}\right]$ *is submodular. Hence*

$$(\delta, s) \rightarrow \sum_{j \in \mathcal{J}} \delta_j s_j - \mathbb{E}_P\left[\max_{k \in J_0}\{\delta_k + \varepsilon_k\}\right]$$

is supermodular in δ *and has increasing differences in* (δ, s). *As a result of Topkis' theorem, the set-valued map* $s \rightarrow \tilde{\Sigma}^{-1}(s)$ *is increasing in the strong set order, which means that if* $s \leq s'$, $\delta \in \tilde{\Sigma}^{-1}(s)$ *and* $\delta' \in \tilde{\Sigma}^{-1}(s')$, *then*

$$\delta \wedge \delta' \in \tilde{\Sigma}^{-1}(s) \ \text{ and } \delta \vee \delta' \in \tilde{\Sigma}^{-1}(s'),$$

which exactly recovers the conclusion of theorem 2. However, as soon as the model is no longer an additive random utility model, $\tilde{\Sigma}^{-1}(s)$ *is not obtained by the solution of a maximization problem, so that Topkis' theorem cannot be invoked.* □

It follows from theorem 2' that the identified utility set is a lattice whenever nonempty, which implies corollary 1. Indeed, if $\delta \in \tilde{\Sigma}^{-1}(s)$ and $\delta' \in \tilde{\Sigma}^{-1}(s)$, then $\delta \wedge \delta' \in \tilde{\Sigma}^{-1}(s)$ and $\delta \vee \delta' \in \tilde{\Sigma}^{-1}(s)$, QED.

A.2.3. *Proof of theorem 3.* As before, theorem 3 can be proven without Assumption 2, which leads us to formulate a result which will imply theorem 3. Define the domain of $\tilde{\Sigma}^{-1}$ as

$$\mathcal{S}_0^{dom} = \left\{ s \in \mathcal{S}_0 : \tilde{\Sigma}^{-1}(s) \neq \emptyset \right\}.$$

We have:

Theorem 3'. *Under Assumptions 1, 3, and 4, $\tilde{\Sigma}^{-1}(s)$ is non-empty for all $s \in \mathcal{S}_0^{int}$.*

The proof is based on the additional lemmas 1-4, which we first state and prove.

Assumption 4 implies that $\forall\, \eta > 0,\ \nu > 0$, there is δ^* s.t. $\delta > \delta^*$ implies $\Pr(|X_\delta - X^*_\delta| > \nu) < \eta$.

Lemma 1. *There is a T^* such that for $T < T^*$ there exists $\underline{\delta}_j^T$ such that*

$$\int \frac{\exp\left(\frac{\mathcal{U}_{\varepsilon j}(\underline{\delta}_j)}{T}\right)}{1 + \exp\left(\frac{\mathcal{U}_{\varepsilon j}(\underline{\delta}_j)}{T}\right)} P(d\varepsilon) = s_j \tag{A.17}$$

and for all $T < T^*$, $\underline{\delta}_j^T \geq \underline{\delta}_j$ where $\underline{\delta}_j$ does not depend on T.

Proof. For $T > 0$, let

$$F_j^T(\delta) = \int \frac{\exp\left(\frac{\mathcal{U}_{\varepsilon j}(\delta)}{T}\right)}{1 + \exp\left(\frac{\mathcal{U}_{\varepsilon j}(\delta)}{T}\right)} P(d\varepsilon) = \int \frac{1}{1 + \exp\left(-\frac{\mathcal{U}_{\varepsilon j}(\delta)}{T}\right)} P(d\varepsilon).$$

Assumption 3, part (b) implies:

<u>Fact (a):</u> $F_j^T(.)$ is continuous and strictly increasing.

Next, by Assumption 4, there exists $\underline{\delta}_j$ such that $\delta < \underline{\delta}_j$ implies $\Pr(\mathcal{U}_{\varepsilon j}(\delta) > -a) \leq s_j/2$. Hence, for $\delta < \underline{\delta}_j$

$$\begin{aligned} F_j^T(\delta) &= \int_{\{\mathcal{U}_{\varepsilon j}(\delta) < -a\}} \frac{\exp\left(\frac{\mathcal{U}_{\varepsilon j}(\delta)}{T}\right)}{1 + \exp\left(\frac{\mathcal{U}_{\varepsilon j}(\delta)}{T}\right)} P(d\varepsilon) + \int_{\{\mathcal{U}_{\varepsilon j}(\delta) \geq -a\}} \frac{\exp\left(\frac{\mathcal{U}_{\varepsilon j}(\delta)}{T}\right)}{1 + \exp\left(\frac{\mathcal{U}_{\varepsilon j}(\delta)}{T}\right)} P(d\varepsilon) \\ &\leq \frac{1}{1 + \exp\left(\frac{a}{T}\right)} + s_j/2 \end{aligned}$$

and taking $T^* = a/\log(1/s_j - 1)$ if $\log(1/s_j - 1) > 0$, and $T^* = +\infty$ else, it follows that $T \leq T^*$ implies $\frac{1}{1+\exp\left(\frac{a}{T}\right)} \leq s_j/2$, hence we get to:

Fact (b): for $\delta < \underline{\delta}_j$ and $T \leq T^*$, one has $F_j^T(\delta) < s_j$.

Next, by Assumption 4, there exists δ'_j such that $\delta > \delta'_j$ implies $\Pr(\mathcal{U}_{\varepsilon j}(\delta) > 0) \geq 2s_j$. Then for $\delta > \delta'_j$,

$$F_j^T(\delta) \geq \int_{\{\mathcal{U}_{\varepsilon j}(\delta) > b\}} \frac{\exp\left(\frac{\mathcal{U}_{\varepsilon j}(\delta)}{T}\right)}{1 + \exp\left(\frac{\mathcal{U}_{\varepsilon j}(\delta)}{T}\right)} P(d\varepsilon) \geq \frac{\Pr(\mathcal{U}_{\varepsilon j}(\delta) > b)}{1 + \exp(0)} \geq \frac{2s_j}{2} = s_j.$$

As a result, we get:

Fact (c): for all $T > 0$ and for $\delta > \delta'_j$, $F_j^T(\delta) > s_j$.

By combination of facts (a), (b) and (c), it follows that for $T \leq T^*$, there exists a unique $\underline{\delta}_j^T$ such that $F_j^T\left(\underline{\delta}_j^T\right) = s_j$ and $\underline{\delta}_j^T \leq \delta'_j$, where δ'_j does not depend on $T \leq T^*$. □

Let

$$G_j^T(\delta_j; \delta_{-j}) := \int \frac{P(d\varepsilon)}{\exp\left(-\frac{\mathcal{U}_{\varepsilon j}(\delta_j)}{T}\right) + \sum_{j' \in \mathcal{J}} \exp\left(\frac{\mathcal{U}_{\varepsilon j'}(\delta_{j'}) - \mathcal{U}_{\varepsilon j}(\delta_j)}{T}\right)}.$$

Lemma 2. *For $T < T^*$, if $G_j^T\left(\delta_j^{T,k}; \delta_{-j}^{T,k}\right) \leq s_j$, then:*

(i) there is a real $\delta_j^{T,k+1} \geq \delta_j^{T,k}$ such that

$$G_j^T\left(\delta_j^{T,k+1}; \delta_{-j}^{T,k}\right) = s_j,$$

(ii) one has $G_j^T\left(\delta_j^{T,k+1}; \delta_{-j}^{T,k+1}\right) \leq s_j$.

Proof. Take $\eta > 0$ such that $\eta < 1 - \sqrt{s_j}$. There is $M > 0$ such that

$$\Pr\left(1 + \sum_{j' \neq j} \exp\left(\frac{\mathcal{U}_{\varepsilon j'}\left(\delta_{j'}^{T,k}\right)}{T}\right) < M\right) > 1 - \eta/2.$$

We have

$$\begin{aligned} G_j^T\left(\delta_j; \delta_{-j}^{T,k}\right) &\geq \int \frac{1\left\{1 + \sum_{j' \neq j} \exp\left(\frac{\mathcal{U}_{\varepsilon j'}\left(\delta_{j'}^{T,k}\right)}{T}\right) < M\right\} P(d\varepsilon)}{1 + \exp\left(-\frac{\mathcal{U}_{\varepsilon j}(\delta_j)}{T}\right)\left(1 + \sum_{j' \neq j} \exp\left(\frac{\mathcal{U}_{\varepsilon j'}\left(\delta_{j'}^{T,k}\right)}{T}\right)\right)} \\ &\geq \int \frac{1\left\{1 + \sum_{j' \neq j} \exp\left(\frac{\mathcal{U}_{\varepsilon j'}\left(\delta_{j'}^{T,k}\right)}{T}\right) < M\right\} P(d\varepsilon)}{1 + \exp\left(-\frac{\mathcal{U}_{\varepsilon j}(\delta_j)}{T}\right) M} \end{aligned}$$

Next, by Assumption 4, for all $b \in \mathbb{R}$, there exists δ_j^* such that $\delta > \delta_j^*$ implies $\Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta\right) > b\right) \geq 1 - \eta/2$. Thus for $\delta_j > \delta_j^*$,

$$G_j^T\left(\delta_j; \delta_{-j}^{T,k}\right) \geq \int \frac{1\left\{1 + \sum_{j' \neq j} \exp\left(\frac{\mathcal{U}_{\varepsilon j'}\left(\delta_{j'}^{T,k}\right)}{T}\right) < M\right\} 1\left\{\mathcal{U}_{\varepsilon j}\left(\delta_j\right) > b\right\} P\left(d\varepsilon\right)}{1 + \exp\left(-\frac{b}{T}\right) M} \geq \frac{1 - \eta}{1 + \exp\left(-\frac{b}{T}\right) M}.$$

Choosing $b = -T\log\left(\eta/M\right)$ implies that the right hand-side is $\frac{1-\eta}{1+\eta} \geq \left(1 - \eta\right)^2$. Because $\eta < 1 - \sqrt{s_j}$, $\left(1 - \eta\right)^2 > s_j$, and therefore for $\delta_j > \delta_j^*$, $G_j^T\left(\delta_j; \delta_{-j}^{T,k}\right) > s_j$. Hence, because $G_j^T\left(\delta_j^{T,k}; \delta_{-j}^{T,k}\right) \leq s_j$, by continuity of $G_j^T\left(.; \delta_{-j}^{T,k}\right)$, there exists $\delta_j^{T,k+1} \in \left(\delta_j^{T,k}, \delta_j^*\right)$ such that

$$G_j^T\left(\delta_j^{T,k+1}; \delta_{-j}^{T,k}\right) = s_j,$$

which shows claim (i). To show the second claim, let us note that $G_j^T\left(\delta\right)$ is decreasing with respect to $\delta_{j'}$ for any $j' \neq j$. Indeed,

$$G_j^T\left(\delta_j; \delta_{-j}\right) = \int \frac{P\left(d\varepsilon\right)}{\exp\left(-\frac{\mathcal{U}_{\varepsilon j}\left(\delta_j\right)}{T}\right) + 1 + \sum_{j' \neq j} \exp\left(\frac{\mathcal{U}_{\varepsilon j'}\left(\delta_{j'}\right) - \mathcal{U}_{\varepsilon j}\left(\delta_j\right)}{T}\right)}$$

is expressed as the expectation of a term which is decreasing in $\delta_{j'}$. Hence, as $\delta_{-j}^{T,k} \leq \delta_{-j}^{T,k+1}$ in the componentwise order, it follows that

$$G_j^T\left(\delta_j^{T,k+1}; \delta_{-j}^{T,k+1}\right) \leq G_j^T\left(\delta_j^{T,k+1}; \delta_{-j}^{T,k}\right) = s_j,$$

which shows claim (ii). □

Because of lemma 2, one can construct recursively a sequence $\left(\delta_j^{T,k}\right)$ such that $\delta_j^{T,k+1} \geq \delta_j^{T,k}$ and

$$G_j^T\left(\delta_j^{T,k}\right) \leq s_j, \tag{A.18}$$

From Assumption 4, setting $\eta = s_0/4$ and $b = T^*\log\left(4/s_0 - 1\right)$, one has the existence of $\delta \in \mathbb{R}$ such that $\delta \geq \overline{\delta}_j$ implies $\Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta\right) < b\right) < \eta$.

Lemma 3. *For all $k \in \mathbb{N}$ and $T < T^*$, one has*

$$\delta_j^k \leq \overline{\delta}_j \tag{A.19}$$

where $\overline{\delta}_j$ is a constant independent from $T < T^*$.

Proof. By summation of inequality (A.18) over $j \in \mathcal{J}$, one has

$$\begin{aligned} s_0 &\leq \int \frac{P\left(d\varepsilon\right)}{1+\sum_{j'\in\mathcal{J}}\exp\left(\mathcal{U}_{\varepsilon j'}\left(\delta_{j'}^{T,k}\right)/T\right)} \leq \int \frac{P\left(d\varepsilon\right)}{1+\exp\left(\mathcal{U}_{\varepsilon j}\left(\delta_{j}^{T,k}\right)/T\right)} \\ &\leq \Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta_{j}^{T,k}\right)<b\right)+\int_{\left\{\mathcal{U}_{\varepsilon j}\left(\delta_{j}^{T,k}\right)\geq b\right\}} \frac{P\left(d\varepsilon\right)}{1+\exp\left(\mathcal{U}_{\varepsilon j}\left(\delta_{j}^{T,k}\right)/T\right)} \\ &\leq \Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta_{j}^{T,k}\right)<b\right)+\frac{1}{1+\exp\left(b/T^*\right)}. \end{aligned}$$

Now assume by contradiction that $\delta_j^{T,k} > \overline{\delta}_j$. Then $\Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta\right)<b\right) < \eta = s_0/4$ and $\left(1+\exp\left(b/T^*\right)\right)^{-1} = s_0/4$, and thus one would have

$$s_0 \leq s_0/4 + s_0/4 = s_0/2,$$

a contradiction. Thus inequality (A.19) holds. □

Lemma 4. *Let* $\delta_j^T = \lim_{k\to+\infty}\delta_j^{T,k}$. *One has*

$$G_j^T\left(\delta_j^T;\delta_{-j}^T\right) = s_j. \tag{A.20}$$

Proof. One has $G_j^T\left(\delta_j^{T,k+1};\delta_{-j}^{T,k}\right) = s_j$; by the fact that $\delta_j^{T,k+1}\to\delta_j^T$ and $\delta_{-j}^{T,k}\to\delta_{-j}^T$ and by the continuity of G_j^T, it follows (A.20). □

We can now deduce the proof of theorem .

Proof of theorem 3'. Point (a): lemma 4 implies that one can define

$$u^T\left(\varepsilon\right) = T\log\left(1+\sum_{j\in\mathcal{J}}\exp\left(\mathcal{U}_{\varepsilon j}\left(\delta_j^T\right)/T\right)\right) \text{ and } \pi_{\varepsilon j}^T = \exp\left(\frac{-u^T\left(\varepsilon\right)+\mathcal{U}_{\varepsilon j}\left(\delta_j^T\right)}{T}\right),$$

and by the same result, one has

$$\mathbb{E}_{\pi^T}\left[u^T\left(\varepsilon\right)\right] = \mathbb{E}_{\pi^T}\left[\mathcal{U}_{\varepsilon j}\left(\delta_j^T\right)\right].$$

It follows from lemma 3 that the sequence δ_j^T is bounded independently of T, so by compactness, it converge up to subsequence toward δ_j^0. Note that $\underline{\delta}_j^0 \leq \delta_j^0 \leq \overline{\delta}_j^0$. We can extract a converging subsequence π^{T_n} where $T_n\to 0$ and $\pi^{T_n}\to\pi^0$ in the weak convergence. Mimicking the argument in Villani (2003) page 32, it follows that $\pi^0\in\mathcal{M}\left(P,s\right)$.

Point (b): Let $u^0\left(\varepsilon\right) = \max_{j\in\mathcal{J}_0}\left\{\mathcal{U}_{\varepsilon j}\left(\delta_j^0\right)\right\}$. We have $u^0\left(\varepsilon\right)\geq\mathcal{U}_{\varepsilon j}\left(\delta_j^0\right)$. Let us show that

$$\mathbb{E}_{\pi^0}\left[u^0\left(\varepsilon\right)\right] = \mathbb{E}_{\pi^0}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right],$$

which will proof the final result. We have $\mathbb{E}_{\pi^{T_n}}\left[u^{T_n}\left(\varepsilon\right)\right] = \mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon j}\left(\delta_j^{T_n}\right)\right]$; let us show that

(i) $\mathbb{E}_{\pi^{T_n}}\left[u^{T_n}\left(\varepsilon\right)\right]\to\mathbb{E}_{\pi^0}\left[u^0\left(\varepsilon\right)\right]$, and
(ii) $\mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^{T_n}\right)\right]\to\mathbb{E}_{\pi^0}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right]$

Start by showing point (i). We have $0 \le u^0(\varepsilon) - u^T(\varepsilon) \le T_n \log \mathcal{J}$. As a result, $\mathbb{E}_{\pi^{T_n}}\left[u^{T_n}(\varepsilon)\right] = \mathbb{E}_P\left[u^{T_n}(\varepsilon)\right] \to \mathbb{E}_P\left[u^0(\varepsilon)\right] = \mathbb{E}_{\pi^0}\left[u^0(\varepsilon)\right]$.

Next, we show point (ii). One has,

$$\mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^{T_n}\right)\right] - \mathbb{E}_{\pi^0}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right] = \mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_j^{T_n}\right) - \mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right] + \mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right] - \mathbb{E}_{\pi^0}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right]$$

Let $\nu > 0$. For any $K \subseteq \mathcal{X}$ compact subset of $\mathcal{X}$, one has

$$\left|\mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^{T_n}\right) - \mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right]\right| \le \mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^{T_n}\right) - \mathcal{U}_{\varepsilon J}\left(\delta_J^0\right) 1_{\{\varepsilon \in K\}}\right] + 2\mathbb{E}_P\left[\sum_j \left|\mathcal{U}_{\varepsilon j}\left(\overline{\delta}_j\right)\right| 1\{\varepsilon \in K\}\right]$$

hence, one may choose K such that $\mathbb{E}_P\left[\sum_j \left|\mathcal{U}_{\varepsilon j}\left(\overline{\delta}_j\right)\right| 1\{\varepsilon \in K\}\right] < \nu/4$. By uniform continuity of $\varepsilon \to \mathcal{U}_{\varepsilon j}(\delta)$ on K, and because $\delta_j^{T_n} \to \delta_j^0$, there exists $n' \in \mathbb{N}$ such that $n \ge \overline{n}$ implies $\max_{j \in \mathcal{J}}\left|\mathcal{U}_{\varepsilon J}\left(\delta_j^{T_n}\right) - \mathcal{U}_{\varepsilon J}\left(\delta_j^0\right)\right| \le \nu/2$ for each $\varepsilon \in K$. Thus, for $n \ge n'$, one has

$$\left|\mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^{T_n}\right) - \mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right]\right| \le \nu \tag{A.21}$$

By the weak convergence of π^{T_n} toward π^0, there is $n'' \ge n'$ such that for $n \ge n''$ one has

$$\left|\mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right] - \mathbb{E}_{\pi^0}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right]\right| \le \nu. \tag{A.22}$$

Combining (A.21) and (A.22), it follows that for $n \ge n''$,

$$\left|\mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^{T_n}\right) - \mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right]\right| \le 2\nu,$$

which establishes point (ii). The result is proven by noting that $\mathbb{E}_{\pi^0}\left[u^0(\varepsilon)\right] = \mathbb{E}_{\pi^0}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right]$ along with $u^0(\varepsilon) \ge \mathcal{U}_{\varepsilon j}\left(\delta_j^0\right)$ for all ε and j implies that $(\varepsilon, j) \in Supp\left(\pi^0\right)$ implies $u^0(\varepsilon) = \mathcal{U}_{\varepsilon j}\left(\delta_j^0\right)$, QED. □

A.2.4. *Proof of theorem 4.* Once again we shall prove this theorem without Assumption 2.

Theorem 4'. *Under Assumptions 1 and 5, $\tilde{\Sigma}^{-1}(s)$ has a single element for all $s \in \mathcal{S}_0^{int}$.*

Proof. Let $\delta \in \tilde{\Sigma}^{-1}(s)$. Define Z to be the random vector such that $Z_j = \mathcal{U}_{\varepsilon j}^{-1}\left(\mathcal{U}_{\varepsilon 0}(\delta_0)\right)$, and P_Z to be the probability distribution of Z. By proposition 1, this implies

$$P\left(U_{\varepsilon 0}(\delta_0) > \max_{j \in \mathcal{J}} U_{\varepsilon j}(\delta_j)\right) \le s_0 \le P\left(U_{\varepsilon 0}(\delta_0) \ge \max_{j \in \mathcal{J}} U_{\varepsilon j}(\delta_j)\right)$$

which is equivalent to

$$P(Z_j > \delta_j) \le s_0 \le P(Z_j \ge \delta_j)$$

but because Z has a density, the latter condition is equivalent to

$$s_0 = P(Z_j \ge \delta_j).$$

Now consider $\delta^{\min}$ and $\delta^{\max}$ the lattice bounds of $\tilde{\Sigma}^{-1}(s)$. Because of the previous remark, $P\left(Z_j \ge \delta_j^{\min}\right) = P\left(Z_j \ge \delta_j^{\max}\right)$. However, as distribution of Z has full support, the map $\delta \to$

$P(Z_j \geq \delta_j)$ is strictly increasing in each δ_j, and as a result of $\delta_j^{\min} \leq \delta_j^{\max}$, it follows that $\delta_j^{\min} = \delta_j^{\max}$, QED. □

A.3. **Proof of theorem 5.** We start with a series of auxiliary lemmas. Because we need to work with different distributions P of ε, we shall in this paragraph make the dependence of $\tilde{\Sigma}^{-1}$ in P and s explicit by writing $\tilde{\Sigma}^{-1}(P,s)$ instead of $\tilde{\Sigma}^{-1}(s)$ as in the rest of the paper.

Lemma 5. *The lattice upper bound $\bar{\delta}$ of $\tilde{\Sigma}^{-1}(P;s)$ is such that*

$$\bar{\delta}_j = \max_{(\delta_{-j}) \in \mathbb{R}^{\mathcal{J} \setminus \{j\}}} F(\delta_{-j}; P, s) \tag{A.23}$$

where $F(\delta_{-j}; P, s) = \min_{B \subseteq \mathcal{J}_0 \setminus \{j\}} F_B^{-1}\left(\sum_j s_j; P, \delta_{-j}\right)$, and $F_{Bj}^{-1}(.; P, \delta_{-j})$ is the generalized inverse of the nonincreasing and left-continuous map defined by

$$F_{Bj}(\delta_j; P, \delta_{-j}) = P\left(\max_{k \in B} U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k) \geq \max_{k \in \mathcal{J}_0 \setminus (B \cup \{j\})} \left\{U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k), \delta_j\right\}\right). \tag{A.24}$$

Proof. By proposition 1, and theorem 2, if $\bar{\delta} = \sup \tilde{\Sigma}^{-1}(P;s)$, then

$$\bar{\delta}_j = \max_{\delta} \{\delta_j\}$$

subject to $\delta \in \mathbb{R}^{\mathcal{J}}$ and for all $B \subseteq \mathcal{J}_0$ such that $j \notin B$

$$P\left(\max_{k \in B} U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k) > \max_{k \in \mathcal{J}_0 \setminus (B \cup \{i\})} \left\{U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k), \delta_j\right\}\right) \leq \sum_{k \in B} s_k \leq F_{Bj}(\delta_j; P, \delta_{-j}).$$

But because the left and right-hand side terms are both nonincreasing in δ_j, and because on seeks the maximum such δ_j, one may discard the left-hand side inequality, and lemma 5 follows. □

Lemma 6. *There is a constant ρ (which does not depend on n) such that $\delta \in \tilde{\Sigma}^{-1}(P^n; s^n)$ implies $\|\delta\| \leq \rho$ almost surely.*

Proof. $\delta \in \tilde{\Sigma}^{-1}(P^n; s^n)$ implies that for all $B \subseteq \mathcal{J}_0$ such that $j \notin B$

$$P^n\left(\delta_j > \max_{k \in \mathcal{J}_0 \setminus \{j\}} \left\{U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k), \delta_j\right\}\right) \leq s_j^n \leq P^n\left(\delta_j \geq \max_{k \in \mathcal{J}_0 \setminus \{j\}} \left\{U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k), \delta_j\right\}\right)$$

to hold for all j, from which a uniform bound on $|\delta_j|$ can be deduced. □

Lemma 7. *There is a constant A which does not depend on n or δ such that for all δ and δ' such that $\max(\|\delta\|, \|\delta'\|) \leq \rho$, one has*

$$\left|F_{Bj}\left(\delta_j'; P, \delta_{-j}\right) - F_{Bj}(\delta_j; P, \delta_{-j})\right| \geq A \left|\delta_j' - \delta_j\right|. \tag{A.25}$$

Proof. Let $X_B(\delta_{-j}) = \max_{k \in B} U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k)$ and $Y_B(\delta_{-j}) = \max_{k \notin B \cup \{j\}} U_{\varepsilon i}^{-1} U_{\varepsilon k}(\delta_k)$. By Assumption 6, $(X_B(\delta_{-j}), Y_B(\delta_{-j}))$ has a nonvanishing continuous density $f_{X_B(\delta_{-j}), Y_B(\delta_{-j})}(x, y; \delta_{-j})$ which depends continuously on δ_{-j}. One has $F_{Bj}'(\delta_j; P, \delta_{-j}) = -\int_{-\infty}^{\delta_j} f_{X_B(\delta_{-j}), Y_B(\delta_{-j})}(\delta_j, y; \delta_{-j})\, dy$. As this term is a function of $\delta \in \mathbb{R}^{\mathcal{J}}$ which is continuous on the ball of radius ρ around 0, a compact set, and is negative on that set, there is some constant $A > 0$ such that $F'_{Bj}(\delta_j; P, \delta_{-j}) < -A$. As a result, inequality (A.25) holds uniformly. $\square$

We are now ready for the proof of the theorem.

Proof of theorem 5. Because Assumption 6 implies the absence of indifference, $\delta \in \tilde{\Sigma}^{-1}(P; s)$ implies that for all $B \subseteq \mathcal{J}_0$ such that $j \notin B$

$$\sum_{k \in B} s_k = F_{Bj}(\delta_j; \delta_{-j}),$$

while $\delta^n \in \tilde{\Sigma}^{-1}(P^n; s^n)$ implies

$$E_{Bj}\left(\delta_j^n; P^n, \delta_{-j}^n\right) \leq \sum_{k \in B} s_k^n \leq F_{Bj}\left(\delta_j^n; P^n, \delta_{-j}^n\right),$$

where F_{Bj} is defined in (A.24), and E_{Bj} is defined by

$$F_{Bj}(\delta_j; P, \delta_{-j}) = P\left(\max_{k \in B} U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k) > \max_{k \in \mathcal{J}_0 \backslash (B \cup \{j\})} \left\{U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k), \delta_j\right\}\right).$$

hence

$$E_{Bj}\left(\delta_j^n; P, \delta_{-j}^n\right) - \zeta_1^n - \zeta_3^n \leq \sum_{k \in B} s_k \leq F_{Bj}\left(\delta_j^n; P, \delta_{-j}^n\right) + \zeta_2^n + \zeta_3^n$$

where

$$\begin{cases} \zeta_1^n = \sup_{\|\delta\| \leq \rho} |E_{Bj}(\delta_j; P^n, \delta_{-j}) - E_{Bj}(\delta_j; P, \delta_{-j})| \\ \zeta_2^n = \sup_{\|\delta\| \leq \rho} |F_{Bj}(\delta_j; P^n, \delta_{-j}) - F_{Bj}(\delta_j; P, \delta_{-j})| \\ \zeta_3^n = \sum_{k \in B} s_k^n - \sum_{k \in B} s_k \end{cases},$$

hence $\eta_n := |\zeta_3^n| + \max(|\zeta_1^n|, |\zeta_2^n|) \geq |F_B(\delta_i^n; \delta_{-i}) - F_B(\delta_i; \delta_{-i})| \geq A|\delta_i^n - \delta_i|$, and thus $|\delta_i^n - \delta_i| \leq \eta_n / A$, QED. $\square$

A.4. **Additional results.** For $s \in \mathcal{S}_0^{dom}$ and $j \in \mathcal{J}$, define

$$\tilde{\delta}_j^{\min}(s) = \min\left\{\delta_j : \delta \in \tilde{\Sigma}^{-1}(s)\right\} \text{ and } \tilde{\delta}_j^{\max}(s) = \max\left\{\delta_j : \delta \in \tilde{\Sigma}^{-1}(s)\right\}.$$

We have the following result, which follows directly from theorem 2':

Proposition 2. *Under Assumption 1, let $s \in \mathcal{S}_0^{dom}$. Then the following holds:*

(i) The set $\tilde{\Sigma}^{-1}(s)$ has a minimal and a maximal element, namely,

$$\tilde{\delta}^{\min}(s) \in \tilde{\Sigma}^{-1}(s) \text{ and } \tilde{\delta}^{\max}(s) \in \tilde{\Sigma}^{-1}(s).$$

(ii) Any $\delta \in \tilde{\Sigma}^{-1}(s)$ is such that

$$\tilde{\delta}^{\min}(s) \leq \delta \leq \tilde{\delta}^{\max}(s).$$

(iii) $\tilde{\Sigma}^{-1}(s)$ *is point-identified if and only if*

$$\tilde{\delta}^{\min}(s) = \tilde{\delta}^{\max}(s).$$

This result is reminiscent of the result by Berry, Gandhi and Haile (2013) on the inverse isotonicity of the demand map under connected substitutes. However, neither result implies the other one, and we now discuss the connection, between them based on the following simple consequence of theorem 2':

Proposition 3. *Under Assumption 1, the following holds:*

(i) Let $s \in \tilde{\Sigma}(\delta)$*;* $s' \in \tilde{\Sigma}(\delta')$ *be such that* $s_j \leq s'_j$ *for all* $j \in \mathcal{J}$*. Then*

$$\delta_j \leq \tilde{\delta}^{\max}_j(s') \text{ and } \delta'_j \geq \tilde{\delta}^{\min}_j(s).$$

(ii) Let $s, s' \in \mathcal{S}_0^{dom}$ *such that* $s_j \leq s'_j$ *for all* $j \in \mathcal{J}$*. Then*

$$\tilde{\delta}^{\min}_j(s) \leq \tilde{\delta}^{\min}_j(s') \text{ and } \tilde{\delta}^{\max}_j(s) \leq \tilde{\delta}^{\max}_j(s')$$

hold for all $j \in \mathcal{J}$*.*

Proof. Let $\delta \in \tilde{\Sigma}^{-1}(s)$. By proposition 2, point (i), $\delta' := \tilde{\delta}^{\max}(s') \in \tilde{\Sigma}^{-1}(s')$. By theorem 2', it follows that $\delta \vee \delta' \in \tilde{\Sigma}^{-1}(s')$. Hence, by proposition 2, point (ii), it follows that $\delta \vee \delta' \leq \delta'$, thus $\delta \leq \delta'$. The other inequality is proven similarly. □

Proposition 3 relates to Theorem 1 of Berry, Gandhi, and Haile (2013). Indeed, these authors show that, under the assumptions that $\tilde{\Sigma}(\delta) = \{\tilde{\sigma}(\delta)\}$ is point-valued, defined on a Cartesian product, satisfies weak substitutes (i.e. $\tilde{\sigma}_j$ is nonincreasing in δ_k for every $j \in \mathcal{J}_0$ and $k \in \mathcal{J}$) and a connected strong substitutes assumption, then $\tilde{\sigma}$ is inverse isotone. It implies that $\tilde{\delta}^{\min}(s) = \tilde{\delta}^{\max}(s)$ and that the function $\tilde{\delta}(s)$ is inverse isotone. In contrast, in our setting, both $\tilde{\Sigma}(\delta)$ and $\tilde{\Sigma}^{-1}(s)$ may not be point valued, which means that $\tilde{\delta}^{\min}(s)$ and $\tilde{\delta}^{\max}(s)$ may differ. But proposition 3 shows that both these lattice bounds are isotone. In the case they coincide, one recovers the same conclusion as Berry, Gandhi, and Haile (2013).

A.5. **Market Shares Adjustment: Algorithm for the lower bound.** In order to calculate the lower bound, one could implement the same algorithm as the one for the upper bound, but invert the roles of yogurts and consumers as in Kelso and Crawford (1982). However, this would be inefficient as there are few brands of yogurts but many different consumers. Therefore, the algorithm would be fast for the upper bound, as it deals with only few δ_j's, but not for the lower bound as it deals with a lot of different u_ε. Instead of switching the roles of consumers and yogurts, we adapt the upper bound algorithm described in section 4 for the lower bound.

We set the initial systematic utility δ_j^{ub} equal to the lattice upper bound (estimated using the algorithm for the upper bound). In the "first loop" below, we iterate from $\left\{\delta_j^{ub}\right\}$ down to values of δ which are below the lower bound $\underline{\delta}$ (corresponding to a vector of δ at which all the non-reference brands $j \neq 0$ are in excess supply and the reference brand 0 is in excess demand). In the "second loop", we iterate up from this point up to the lower bound, similarly to the MSA upper bound algorithm.

Algorithm 3 (MSA lower bound)**.** *Take* $\eta^{init} = 1$, $\delta_j^{init} = \delta_j^{ub}$ *and* $block_j = 0$ *for all* $j \in \mathcal{J}$*.*

<u>*Begin first loop*</u>

Require $\left(\delta_j^{init}, \eta^{init}, block_j\right)$.

Set $\eta = \eta^{init}$ *and* $\delta^0 = \delta^{init}$.

While $\eta \geq \eta^{tol}$

Set $\pi_{ij} = 1$ *if* $j \in \arg\max_j \mathcal{U}_{\varepsilon j}(\delta_j)$, *and* $= 0$ *otherwise (breaking ties arbitrarily).*

If $\sum_j block_j = |\mathcal{J}|$, *then set* $\delta_j \leftarrow \delta_j + 2\eta$ *and* $block_j \leftarrow 0$ *for all* $j \in \mathcal{J}$, *and* $\eta \leftarrow \eta/4$.

Else set

$$\delta_j \leftarrow \delta_j - \eta \mathbb{1}\left\{\sum_i \pi_{ij} \geq m_j\right\} \text{ for all } j \in \mathcal{J}$$

If $\mathbb{1}\left\{\sum_i \pi_{ij} < m_j\right\}$, *then* $block_j \leftarrow \mathbb{1}\left\{\sum_i \pi_{i0} > m_0\right\}$

Else $block_j \leftarrow block_j$

End While

Return $\delta^{return} = \delta$.

<u>*End first loop*</u>

<u>*Begin main second loop*</u>

Take $\eta = \eta^{tol}$ *and* $\delta_j^{init} = \delta_j^{return}$.

Repeat:

Call the inner second loop *with parameter values* $\left(\delta_j^{init}\right)$ *which returns* $\left(\delta^{return}\right)$.

Set $\delta^{init} \leftarrow \delta^{return} - 2\eta^{tol}$

Until $\delta_j^{return} > \delta_j^{init}$ *for all* $j \in \mathcal{J}$.

<u>*End main second loop*</u>

Begin inner second loop

Require $\left(\delta_j^{init}\right)$.

Set $\delta = \delta^{init}$.

Set $\pi_{i0} = 1$ *if* $0 \in \arg\max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j)$, *and* $= 0$ *otherwise (breaking ties arbitrarily).*

While $\sum_i \pi_{i0} > m_0$

Set $\pi_{ij} = 1$ *if* $j \in \arg\max_j \mathcal{U}_{\varepsilon j}(\delta_j)$, *and* $= 0$ *oherwise (breaking ties arbitrarily).*

Set $\delta_j \leftarrow \delta_j + \eta \mathbb{1}\left\{\sum_i \pi_{ij} < m_j\right\}$ *for all* $j \in \mathcal{J}$.

Set $\pi_{i0} = 1$ *if* $0 \in \arg\max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j)$, *and* $= 0$ *otherwise (breaking ties arbitrarily).*

End While

Return $\delta^{return} = \delta$.

End inner second loop

Appendix B. Semi-discrete transport algorithms

In this section we describe an additional computational algorithm which is specialized for solving the particular case of the pure characteristics demand model (discussed as Example 2.1 above). The method discussed here can be used when the distribution of the unobserved taste vector is uniformly distributed over a polyhedron, typically the Cartesian product of compact intervals. Recall that the pure characteristics model has $\varepsilon_{ij} = \nu_i^\intercal x_j$, with $\nu \sim \mathbf{P}_\nu$ is a random vector distributed over $\mathbb{R}^d$, and assume that $\mathbf{P}_\nu$ is the uniform distribution over $E = \prod_{1 \leq k \leq d} [0, l_k]$. Then we can use the equivalence theorem in order to compute $\tilde{\sigma}^{-1}$ using semi-discrete transport algorithms, which were pioneered by Aurenhammer (1987), with substantial progress made recently by Kitagawa, Mérigot, and Thibert (2016) and Lévy (2015). The idea, exposited in chapter 5 of Galichon (2016), is that the optimal transport problem (3.6) can be reformulated as a finite-dimensional unconstrained convex optimization problem

$$\inf_{\delta \in \mathbb{R}^{\mathcal{J}_0}} F(\delta), \text{ where } F(\delta) = \mathbb{E}_P\left[\max_{j \in \mathcal{J}_0} \{\delta_j + \varepsilon_j\}\right] - \sum_{j \in \mathcal{J}_0} \delta_j s_j. \tag{B.1}$$

Semi-discrete algorithms consist of a gradient descent over F. Note that $\partial F / \partial \delta_j = \Pr\left(\forall j' \in \mathcal{J}_0 \backslash \{j\}, \varepsilon_{j'} - \varepsilon_j \leq \delta_j - \delta_{j'}\right) - s_j$, where the first term is the area of the polytope $\left\{\varepsilon \in E : \forall j' \in \mathcal{J}_0 \backslash \{j\}, \varepsilon_{j'} - \varepsilon_j \leq \delta_j - \delta_{j'}\right\}$, hence a gradient descent can be done provided one can compute areas of polytopes. The Hessian of F can be computed relatively easily too; we refer to Kitagawa, Mérigot, and Thibert (2016) for details.

For these semi-discrete approaches, we provide an R-based interface to Geogram by Lévy (2018). Geogram is a C++ library of geometric algorithms, with fast implementations of semi-discrete optimal transport methods with two or three random coefficients. The package is open-source and is available on GitHub as Lévy and O'Hara (2018).

For the Pure Characteristics Model, the semi-discrete algorithm provides super-fast performance, far outstripping all the other algorithms (cf. table 7 below). Moreover, the semi-discrete algorithm does not rely on simulations, and computes the exact solution set of δ_j which ensures equilibrium in the market for given market shares and given vector of observed characteristics x_j. The drawbacks are that, as far as we are aware, the semi-discrete approach is only available for random coefficient distributions ν_i which are jointly-uniform with at most three dimensions; which limits its use in many applications of the pure characteristics models (which typically assumed a joint Gaussian distribution for the random coefficients). In table 7, we compute Monte-Carlo simulations similar to the ones showed in table 2 except that the random tastes shocks ν_i are drawn from independent uniform distributions and not from Gaussian ones.

Table 7. Average computational Time (secs.) for pure characteristics models with uniform random shocks

Algorithms	Draws	Brands	RMSE	CPU time (secs)
BLP contract. map.	1,000	5	0.062	0.034
LP (gurobi)	1,000	5	0.009	0.104
Auction	1,000	5	0.010	0.009
Semi discrete	1,000	5	0	0.003
MSA	1,000	5	0.009	5.422
BLP contract. map.	1,000	50	0.032	0.283
LP (gurobi)	1,000	50	0.004	0.358
Auction	1,000	50	0.004	0.044
Semi discrete	1,000	50	0	0.046
BLP contract. map.	1,000	500	0.011	2.766
LP (gurobi)	1,000	500	0.001	3.392
Auction	1,000	500	0.001	0.480
Semi discrete	1,000	500	0	0.743
BLP contract. map.	10,000	5	0.061	0.331
LP (gurobi)	10,000	5	0.003	0.311
Auction	10,000	5	0.003	0.122
Semi discrete	10,000	5	0	0.003
MSA	10,000	5	0.003	1.471
BLP contract. map.	10,000	50	0.032	2.894
LP (gurobi)	10,000	50	0.001	3.947
Auction	10,000	50	0.001	0.735
Semi discrete	10,000	50	0	0.046
BLP contract. map.	10,000	500	0.011	33.171
LP (gurobi)	10,000	500	0.000	54.824
Auction	10,000	500	0.000	5.451
Semi discrete	10,000	500	0	0.749

Note: The numbers are average of 50 Monte-Carlo replication. Demand inversion for the pure characteristics model with 5, 50 and 500 brands of yogurt and $1,000$ and $10,000$ draws of taste shocks. The column "RMSE" corresponds to the root mean squared of in the estimation of the estimated δ_j. Semi-discrete displays 0 RMSE since there are no sampling errors.

Appendix C. Additional Details of Monte Carlo Simulations of Pure Characteristics Models

In this section, we provide additional details of the data generating process (DGP) and the setup of the tuning parameters for the solvers to reproduce results in Section 5.1. We first introduce our notation. We use the lower-case subscript (e.g., consumer i) to index a generic element in a set (e.g., the set of consumer, $\mathcal{I}$), which is written by the calligraphy font. The cardinality of a set is written by the upper-case letter (e.g., the number of consumer, I). We index consumers by i, product by j, markets by m, (exogenous) product characteristics by k, and instrumental variables by n. We denote by s the market share, by x the product characteristics, and by p the price.

Our DGP for Table 1 is adapted from Dubé, Fox, and Su (2012). We first generate the market- and product-specific regressors $x_{mj} = (x_{mj1}, x_{mj2}, x_{mj3})$ from multivariate normal with

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & -0.7 & 0.3 \\ \cdot & 1 & 0.3 \\ \cdot & \cdot & 1 \end{pmatrix}. \tag{C.1}$$

The unobserved fixed effect ξ_{mj} is independently generated from a normal distribution with mean equals zero and standard deviation equals 1. The price, p_{mj}, is generated according to

$$p_{mj} = |1.1(x_{mj1} + x_{mj2} + x_{mj3}) + 0.5\xi_{mj} + e_{mj}|. \tag{C.2}$$

where e_{mj} is independently generated from a normal distribution with mean equals zero and standard deviation equals 1. The utility of consumer i who chooses alternative j in market m is generated by

$$u_{mij} = \beta_0 - \beta_p p_{mj} + \beta_1 x_{mj1} + \beta_2 x_{mj2} + \beta_3 x_{mj3} + \xi_{mj}. \tag{C.3}$$

We set $\beta_0 = 1$. $(\beta_p, \beta_1, \beta_2, \beta_3)$ are individual-specific coefficients generated from independent normal distributions with the means equal (-1,0.5,0.5,0.2) and the standard deviations all equal 1. To generate the market share data for Table 1, we simulate 1,000 consumers for each market. As the purpose is comparing the numerical performance, following Dubé, Fox, and Su (2012), the same set of simulated consumers are applied to both the data simulation and all estimation algorithms.

Next, we describe the DGP for instrumental variables. We first generate six basis instrument variables z, independently from the following specification:

$$z_{mj} = 0.25\,(1.1\,(x_{mj1} + x_{mj2} + x_{mj3}) + e_{mj}) + u_{mj}, \tag{C.4}$$

where u follows a uniform distribution on the unit interval. We use the linear term of x and z, the quadratic and cubic terms of x and z, the product terms ($\Pi_{k=1}^{3} x_{mjk}$ and $\Pi_{k=1}^{6} z_{mjk}$), and the interaction terms, ($x_{mjl} z_{mjk},\ l = \{1,2\}, k = \{1,\ldots,6\}$). There are total 41 instrumental variables.

Below we provide the setup of tuning parameters for various algorithms involved in Table 1. For the inner loop of Matching-NFXP, we use the combined LP formulation and solved it by GUROBI 9.0. For MPCC and BLP-MPEC, we use KNITRO 12.0. Our choice of the solver depends on the natural of the problem to achieve the best performance. GUROBI is optimized for LP, whereas KNITRO is a general-purpose nonlinear program solver with capability of handling complementarity constraints in Pang, Su, and Lee (2015). KNITRO is also recommended by Dubé, Fox, and Su (2012) for estimating the mixed logit demand. We program all these three algorithms in AMPL and execute them from AMPL's R interface. The Matching-NFXP shares similar computational features as in BLP. As illustrated by Nevo (2000), one essentially minimizes the scale parameters only: Given the the scale parameters, one applies the demand inversion to obtain δ as the "dependent variable" and perform the constrained 2SLS (due to the normalization) to obtain the location parameter estimate. For Model I in Table 1, the problem boils down to a convex programming problem. For Model II in Table 1, we use the **optimize** function in R to find the optimal σ_p. It is based on the golden section search. We set [0.001,5] as the search interval. For MPCC, we deploy the Intel Pardiso MKL in KNITRO using 16 threads. This is to ensure an equal footing since GUROBI used in Matching-NFXP automatically deploys a parallel solver. Since it is extremely costly to run MPCC, we set a low tolerance in KNITRO (xtol and ftol to 1e-04) and use only one starting point. For BLP-MPEC, we use the logit-smoothed AR simulator.[36] to approximate the demand map. One first generate the individual-level simulators for each product characteristics: $(v_{mip}, v_{mi1}, v_{mi2}, v_{mi3})$. The following formulation can be viewed as applying Dubé, Fox, and Su (2012) to Berry and Pakes (2007).

[36]See Train (2009). Berry and Pakes (2007) also use the same method.

$$\begin{aligned}
\min_{\delta,\beta,\sigma} \quad & g^{'}Wg \\
\text{s.t.} \quad & g_n = \textstyle\sum_{m=1}^{M}\sum_{j=1}^{J}(\delta_{mj} - x^{'}_{mj}\beta + \alpha p_{mj})z_{mjn} \\
& \log(s_{mj}) = \log\left(\frac{1}{I}\textstyle\sum_{i=1}^{I}\left(\frac{\exp\left([\delta_{mj} + \sigma_p p_{mk} v_{mip} + \sum_{k=1}^{K}\sigma_k x_{mj} v_{mik}]/\lambda\right)}{\sum_{j'\in\mathcal{J}}\exp\left([\delta_{mj'} + \sigma_p p_{mk} v_{mip} + \sum_{k=1}^{K}\sigma_k x_{mj} v_{mik}]/\lambda\right)}\right)\right)
\end{aligned} \tag{C.5}$$

Notice that the market share constraints are stated in terms of the log of the shares as opposed to Dubé, Fox, and Su (2012). We find this transformation significantly improves both numerical stability and speed; the typical BLP fixed-point iteration is also executed in the log scale. We use 10 starting points, automatically chosen by KNITRO. The smoothing parameter of BLP-MPEC is $\lambda = 1$. Smaller value of λ can increase the accuracy; however, we find that it often results in a steep gradient that cannot be evaluated by KNITRO.

The DGP for Table 2 is described below: The $x_j = (x_{j1}, x_{j2}, x_{j3})$ are drawn from multivariate normal with

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & -0.7 & 0.3 \\ \cdot & 1 & 0.3 \\ \cdot & \cdot & 1 \end{pmatrix}. \tag{C.6}$$

Each consumer i have three tastes shocks ν_i generated from independent normal distributions with means equal $(0.5, 0.5, 0.2)$ and standard deviations all equal 1. Different draws of consumers are used for simulation and estimation thus leading to sampling error in the estimates.