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

ODRAN BONNET[♭], ALFRED GALICHON[†], KEITH O'HARA[◊], AND MATT SHUM[§]

Abstract. We show that the problem of demand inversion – which is a crucial step in the estimation of discrete-choice demand models – is equivalent to the determination of stable outcomes in matching models. This general result applies to random utility models that are not necessarily additive or smooth. Based on this equivalence, algorithms for the determination of stable matchings can provide effective computational methods for estimating these models. This approach permits estimation of models that were previously difficult to estimate, such as the pure characteristics model, as well as non-additive random utility models. We also exploit the theory of stable matchings in order to derive identification results for the utilities solving the demand inversion problem, and to characterize the existence and uniqueness of identified utilities.

Keywords: random utility models, two-sided matching, deferred acceptance, partial identification, pure characteristics model

JEL Classification: C51, C60

Date: October 5, 2018 (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, 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, 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

[◊]New York University; keith.ohara@nyu.edu

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

1. Introduction

Discrete choice models based on random utility maximization have had a tremendous impact on 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 model (RUM) pioneered by McFadden (1978, 1981) 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 characteristics which affects the agent's utility associated with each alternative j. The agent's program is then

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

We assume that 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. Thus we focus throughout on parametric random utility models. Particular instances of these models are additive random utility models (ARUM), such as the 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 setting extends to the more general class of non-additive random utility models (NARUM), in which $\mathcal{U}_{\varepsilon_i j}(\delta_j)$ is no longer necessarily quasilinear in δ_j[1].

Under the assumptions that with probability one, agents are not indifferent between any pair of alternatives, the *demand map* $\sigma(.)$ is defined as the probability that alternative j dominates all the other ones, given the vector of systematic utilities $(\delta_j)_{j \in \mathcal{J}}$. Formally,

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

The main focus of the paper pertains to *demand inversion*: given 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 $s = \sigma(\delta)$? A partial identification issue may arise, as the identified set of vectors δ that solve the demand inversion problem may not necessarily be restricted to a single point.

[1]Typically, frameworks considering time or risk preferences are naturally non additive, see e.g. Apesteguia and Ballester (2014).

To answer this question, 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 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 worker-firm equivalent matching market. Thus, the identified set of solution vectors δ coincides with the set of equilibrium wages in the matching market.

The equivalence between discrete choice and two-sided matching has two practical consequences which allows us to fully address several key issues:

(1) **Computation of the identified set using matching algorithms**. The first benefit of the equivalence principle and reformulation as a two-sided matching problem – a.k.a. transport problem – is that one can use matching algorithms in order to compute the identified set. Fortunately, numerical methods for matching and transport problems are by now well developed. This approach is particularly useful in models that were known to be computationally difficult to estimate, such as the *pure characteristic model*, as well as a wide class of NARUM's. Also, the approach is operative no matter if the underlying model is point- or partially-identified, and it will recover accordingly either the unique or the multiple identified values of δ.

(2) **Characterization of the structure of the identified set δ.** It follows from the equivalence principle that the identified set of vectors δ_j is a lattice. Much of the existing literature on discrete choice models proceeds by introducing conditions which guarantee point identification of δ. Here, our procedures do not require such conditions. Our characterization allows to bypass these assumptions and handle situations of partial identification, which existing methods cannot handle. The lattice structure provides a very simple data-dependent test for point-identification: indeed, if the greatest element of the lattice coincides with its smallest element, then

the set is point-identified.[2] Additionally, we introduce different assumptions which ensure existence, point-identification and consistency.

1.1. **Existing literature.** The simplest setting for demand inversion is the logit model, which is an additive random utility model where the entries of $\varepsilon_i \in \Omega = \mathbb{R}^{\mathcal{J}}$ are interpreted as the utility shocks ε_{ij} associated with each alternative, assumed to be distributed as type I-Extreme Value random variables. There, identification of δ is provided up to an additive constant c by the relation $\delta_j = \log s_j + c$, where c is usually fixed by setting $\delta_0 = 0$ for a particular reference good $j = 0$, yielding the "log-odds ratio" formula,

$$\delta_j - \delta_0 = \log\left(s_j / s_0\right). \tag{1.3}$$

However, the logit model has well-known limitations, and in particular implies very constrained substitution patterns. This has led the literature to investigate more general specifications, for which an explicit inversion formula such as (1.3) does not exist.

Following Berry (1994) and Berry, Levinsohn, and Pakes (1995) many authors have performed demand inversion when the distribution of utility shocks is known, at least up to a parameter. There are several reasons for wishing to recover the full vector of utilities (δ_j) from the observed market shares (s_j). A number of authors have shown the importance of demand inversion approaches to estimating dynamic discrete choice problems à la Rust (1987) and have developed various algorithms based on this inversion (Hotz and Miller (1993), Aguirregabiria and Mira (2002), Arcidiacono and Miller (2011), Dubé, Fox, and Su (2012), Kristensen, Nesheim, and de Paula (2014)). Others, following Berry (1994) have shown the usefulness of this nonparametric inversion approach for solving endogeneity problems with instruments (Berry, Levinsohn, and Pakes (1995),Nevo (2000), Berry and Haile (2014)).

From an analytic perspective, there is a large literature which tackles the question of the invertibility of demand. Berry (1994) shows existence and uniqueness of the vector δ under continuity conditions. Magnac and Thesmar (2002), Norets and Takahashi (2013), and Chiappori, Komunjer, and Kristensen (2009) among others, also tackle this issue under

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

continuity and differentiability assumptions. Berry, Gandhi, and Haile (2013) consider general demand models broader than random utility discrete choice models, and introduce important sufficient conditions for the uniqueness which, unlike previous proposals, do not rely on differentiability assumptions and have a clear economic interpretation. Recently, Chiong, Galichon, and Shum (2016) showed that the problem of demand inversion in the ARUM case is a linear programming problem, and derived new techniques accordingly for computing the utility vector based on the market shares.

In the additive case and when the utility shocks are continuously distributed and have full support, it is well known in the literature that the demand inversion problem has a unique solution. However, outside of these assumptions, the problem of the general characterization of the set of utility vector (δ_j) that are compatible with a vector of market shares (s_j), and the computation of this set, has not been fully resolved in the existing literature to date. 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". This is precisely the gap that our paper sets out to fill. We propose a novel characterization of the set of compatible utility vectors as stable payoffs in a two-sided game with imperfect transferable utility, as studied e.g. in Demange and Gale (1985). We show that the identified set of utility vectors has a lattice structure, and therefore has a greatest and smallest identified element. We show that greatest and smallest utility vectors can be computed by existing algorithms for two-sided matching models, the most well-known of which is the deferred-acceptance algorithm. The underlying mathematical structure at the core of the equivalence between these problems is the notion of a "Galois connection," recently introduced in economics by Nöldeke and Samuelson (2018).

Finally, this paper is also related to papers in the matching literature which point out commonalities between discrete-choice models and aggregate two-sided matching models, with a continuum of agents (Dagsvik (2000), Choo and Siow (2006), Galichon and Salanié (2017), Menzel (2015)).

1.2. **Organization.** Section 2 presents the general random utility framework which is the focus of this paper, and provides examples and comparisons with the more restrictive ARUM framework. Then, section 3 introduces the equivalence of NARUM with two-sided matching problems, reveals the lattice structure of the identified set and tackles the issues of existence, uniqueness and consistency. Section 4 introduces different algorithms which can solve a wide variety of random utility models, along with a discussion of our implementation of these methods, and section 5 contains two numerical investigations of the algorithms, including the pure characteristics model. In the appendix we extend our framework to the case allowing for indifference between two alternatives, and we show that most properties are preserved; in particular, the identified set is still a lattice. All proofs are collected in the appendix.

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 the others will be benchmarked and normalized against (see below section 2.2 about normalization). 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 an 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. The first assumption expresses that the utility derived by an agent from alternative j is an increasing function of the associated systematic utility δ_j.

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.

The second assumption rules out any indifference between two alternatives, or more precisely, expresses that for any vector of systematic utilities δ, there is zero probability that an agent is indifferent between two alternatives.

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',*

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

Assumption 1 ensures that the set of market shares associated with utility vector δ is non-empty, and assumption 2 ensures that it is restricted to a single element. Importantly, however, 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 assumption 2, the market share of alternative j can be defined as the fraction of consumers who prefer weakly *or* strictly alternative j to any other one:

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

Under assumptions 1 and 2, the vector of market shares $s = \sigma\left(\delta\right)$ 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 2, the *demand map* is the map $\sigma : \mathbb{R}^{\mathcal{J}_0} \to \mathcal{S}_0$ defined by expression (2.2).

In appendix A, we will investigate the situation when assumption 2 is violated, implying that agents may be indifferent between available alternatives with a positive probability. As we will see, our main equivalence result holds even without assumption 2, at the cost of using the heavier machinery of set-valued functions, which are discussed in appendix A. In the main text we will maintain assumption 2; as it turns out, this assumption already encompasses models in which the utilities are point or partially identified.

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 case of an additive model where $\mathcal{U}_{\varepsilon_i j}(\delta_j) = \delta_j + \varepsilon_{ij}$, imposing normalization (2.3) is innocuous in the sense that the vector of systematic utilities (δ_j) will yield the same choice problem as the vector $(\delta_j + c)$ where c is a constant. In additive models, any normalization will thus yield the same identified utility vectors δ up to an additive constant. However, this is no longer the case in nonadditive models and the normalization (2.3) will entail loss of generality, which we will explore in our numerical simulations in Section 5.

2.3. **Examples.** Next, we consider several examples of random utility models falling within our framework.

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

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

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 one gets the Logit model, where the demand map is defined as $\sigma_j(\delta) = \exp(\delta_j) / \left(\sum_{j' \in \mathcal{J}_0} \exp(\delta_{j'})\right)$.

Pure characteristics model: In the pure characteristics model, contrary to the logit models, 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

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

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) underlines that this model is appealing and prevalent in theoretic work, but much less in empirical work, arguably due to computational challenges.[3] Our equivalence theorem between discrete choice models and matching models gives us tractable ways to estimate it, which we will highlight in the simulations in section 5.

Random coefficient logit model: In the random coefficient logit model popularized by Berry, Levinsohn, and Pakes (1995) and McFadden and Train (2000), the utility shock is the sum of two independent terms which are each distributed as one of the models above: one logit term ζ_{ij}, and one pure characteristics term $\nu_i^\intercal x_j$, so that the random shock is given by

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

so that T large recovers the logit model, $T \to 0$ the pure characteristics model.

The literature has emphasized several reasons to prefer the pure characteristics to the random coefficient logit model. First, 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.[4] Additionally, Berry, Linton, and Pakes (2004) show that estimators of the pure characteristics model are consistent under weaker assumptions than the random coefficient logit one. Specifically, they prove that the random coefficient logit model will not be consistent if the number of products grows as fast as the number of consumers while the unidimensional pure characteristics model will be.[5]

[3]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.

[4]See Berry and Pakes (2007) and Ackerberg and Rysman (2005)

[5]Let J be the number of products and N the number of consumers. The random coefficient logit model à la BLP requires that $JlogJ/N$ tends to 0 while the unidimensional Pure Characteristics model only requires that N grows at rate $logJ$.

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 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

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

where the expectation is taken over η_j holding ε_i constant. These kind of models are typically non-additive in ε.[6]

Example 2.3 (Vertical differentiation model)**.** We consider an example of the classic vertical differentiation demand framework.[7] Assume that 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 $u + \delta_1$ if retiring (option $j = 1$), in which case she gets pension income y_1. Agent i's non-labour income is ε_i. Then

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

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

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

Example 2.5. Investments with taxes: $\varepsilon \in \{0,1\} \times \mathbb{R}^{\mathcal{J}}$; $\varepsilon^1 = 1$ if tax-exempt individual, $\varepsilon^1 = 0$ if tax-liable; $\delta_j + \varepsilon_j^2$= project j's pre-tax earnings, tax rate $\tau \in (0,1)$, so

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

3. Demand Inversion and the Equivalence Result

We start by working under assumptions 1 and 2, the latter assumption guaranteeing the existence of a demand function σ defined in (2.2), that is the uniqueness of the market shares implied by a given utility vectors δ_j. The demand inversion problem, then, is: given a vector of observed market share $s \in \Delta$, recovering the vector of systematic utilities (δ_j) that satisfy $s_j > 0$ and $\sum_{j \in \mathcal{J}_0} s_j = 1$. Formally,

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

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

The discrete choice problem is traditionally considered a one-sided problem, as apparently consumers choose yogurts more often than the opposite. However, we are now going to demonstrate an equivalence with a two-sided problem: a "marriage problem" between consumers and yogurts, where both sides of the market must assent to be matched. This market will be observationally equivalent to a (hypothetical) dual market where "yogurts choose consumers," hence the title of the paper.

3.1. **The Equivalence Theorem: general case.**

Our central argument in this paper is that the problem of identifying the utilities in the discrete choice problem is equivalent to finding the set of stable outcomes in the matching problem. We introduce a matching game between consumers and yogurts, which is essentially the matching model introduced by Demange and Gale (1985); our presentation of this model is inspired by the one given in chapter 9 of Roth and Sotomayor (1992)[8]. Let $\mathcal{M}(P,s)$ be the set of probability distributions on $\Omega \times \mathcal{J}_0$ 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 measurable subset 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 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 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 $f_{\varepsilon j}(u_\varepsilon) + g_{\varepsilon j}(v_j) = 0$.

The first condition 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 of finding a consumer with utility shock ε matched with a yogurt of type j. As a result, $\pi(j|\varepsilon)$ is the conditional choice probability that an individual with utility shock ε chooses yogurt j. The second condition implies that if an individual with utility shock ε and a yogurt of type j satisfy $f_{\varepsilon j}(u_\varepsilon) + g_{\varepsilon j}(v_j) < 0$, then there exists $u' > u_\varepsilon$ and $v' > v_j$ such that $f_{\varepsilon j}(u') + g_{\varepsilon j}(v') = 0$. Thus there will exist a pair of utilities which is feasible for (ε, j) and which strictly improves upon their equilibrium payoffs u_ε and v_j, which is ruled out in equilibrium. The third condition implies that, if (ε, j) are actually matched, then their

[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.

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 : $f_{\varepsilon j}(u) + g_{\varepsilon j}(v) = 0$.

The following result 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 buyers of yogurts, 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: 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 $\delta_j = -v_j$: the higher v_j, the more demanding a potential partner of type j becomes, making j a less attractive option for any ε, thus, the lower the systematic utility δ_j in the discrete choice model becomes. It is therefore immediate to show that: (i) the no-blocking condition is satisfied: for all ε and j, $u_\varepsilon = \max\limits_{j'\in\mathcal{J}_0} \mathcal{U}_{\varepsilon j'}\left(\delta_{j'}\right) \geq \mathcal{U}_{\varepsilon j}(\delta_j)$; and (ii) the feasibility pair condition is satisfied: if $(\varepsilon, j) \in Supp(\pi)$ then $u_\varepsilon = \max\limits_{j'\in\mathcal{J}_0} \mathcal{U}_{\varepsilon j'}\left(\delta_{j'}\right) = \mathcal{U}_{\varepsilon j}(\delta_j)$.

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. This is discussed in the appendix below and the general result is stated as theorem 1'. Without assumption 2, the equilibrium in the associated matching game may no longer be unique, but the equivalence result in that case implies that the identified set of utilities corresponds to the equilibrium set of the matching model.

3.1.1. *Do consumers choose yogurts or 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 $\boldsymbol{\delta}$, 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 $\delta_j = \min_\epsilon \mathcal{U}^{-1}_{\epsilon j}(u_\epsilon)$. 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" – form a *Galois connection*, a concept that Nöldeke and Samuelson (2018) have also fruitfully applied to mechanism design problems. This Galois property formalizes the notion that in the equilibrium of the two-sided market, matches occur under *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.[9]

3.2. **Additive 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

[9]For obvious reasons, naturally, we prefer the interpretation that consumers choose yogurts!

Salanié (2017), this problem now becomes a matching problem with transferable utility, where the joint surplus of a match between a worker ε and a firm j is ε_j, which is the total match payoff to be split between the salary of the worker and the profit of the firm. This type of problem known as an *optimal transport* problem, a.k.a. Monge-Kantorovich problem, which is a particular class of infinite-dimensional linear programming problems.

It is a well-known result in the optimal transport theory (see e.g. Galichon (2016)) 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

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

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

$$\begin{aligned} & \inf_{u,\delta:\delta_0=0} \left\{\mathbb{E}_P\left[u_\varepsilon\right] - \mathbb{E}_s\left[\delta_{\tilde{j}}\right]\right\} \\ s.t. \qquad & u_\varepsilon - \delta_j \geq \varepsilon_j \ \forall \varepsilon \in \Omega,\ j \in \mathcal{J}_0, \end{aligned} \tag{3.3}$$

which boils down to a 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 will discuss optimal transport algorithms which can be used to perform demand inversion in ARUM models.

3.3. **Properties of the identified set.** As a consequence of the Equivalence theorem, we can employ a number of results from matching theory to describe basic properties of the identified set $\tilde{\sigma}^{-1}(s)$. In particular, we show that the set-valued function $s \to \tilde{\sigma}^{-1}(s)$ is isotone (in a sense to be made precise) and that $\tilde{\sigma}^{-1}(s)$ has a lattice structure (section 3.3.1); we shall provide mild conditions under which it is non-empty (section 3.3.2), conditions under which it is restricted to a point and study its robustness to a perturbation of s and P (section 3.3.3).

3.3.1. *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) 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 the possibility that 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 that of isotonicity of Berry, Gandhi and Haile. As we shall see, the correct generalization is the notion of isotonicity with respect to Veinott's strong set order (see e.g. Veinott (2005)), according to which if $s \leq s'$ and $\delta \in \tilde{\sigma}^{-1}(s)$ and $\delta' \in \tilde{\sigma}^{-1}(s')$, then $\delta \wedge \delta' \in \tilde{\sigma}^{-1}(s)$ and $\delta \vee \delta' \in \tilde{\sigma}^{-1}(s')$, where we recall that the lattice "join" and "meet" operators ($\wedge$ and $\vee$) are defined by $\left(\delta \wedge \delta'\right)_j := \min\left\{\delta_j, \delta'_j\right\}$ and $\left(\delta \vee \delta'\right)_j := \max\left\{\delta_j, \delta'_j\right\}$. When $s \to \tilde{\sigma}^{-1}(s)$ is a point-valued map, we clearly recover the isotonicity of the inverse demand obtained by Berry, Gandhi and Haile.

The main property we deduce from the isotonicity of the inverse demand is the lattice property, which implies that the normalized set of identified utilities has a "largest" (resp. "smallest") element which is composed of the *element-wise* upper- (resp. lower-) bounds among all the utility vectors in the identified set. Practically, most applications of partially identified models focus on computation of the element-wise upper and lower bounds of the identified set of parameters[10]; the lattice result here implies that, for NARUM models, these element-wise upper and lower bounds constitute upper and lower bounds for the parameter vector as a whole. We state:

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 $\delta \vee \delta' \in \tilde{\sigma}^{-1}(s')$.*

As a consequence of theorem 2, we obtain (by taking $s = s'$) that whenever it is nonempty, the set $\tilde{\sigma}^{-1}(s)$ is a *lattice*[11]: if $\delta \in \tilde{\sigma}^{-1}(s)$ and $\delta' \in \tilde{\sigma}^{-1}(s)$, then $\delta \wedge \delta' \in \tilde{\sigma}^{-1}(s)$ and $\delta \vee \delta' \in \tilde{\sigma}^{-1}(s)$. Formally:

[10]Among many others, see Ciliberto and Tamer (2009) and Pakes, Porter, Ho, and Ishii (2015).

[11]Whether the identified set is empty is considered in the next section.

Corollary 1. *Under assumption 1 and 2, if $\tilde{\sigma}^{-1}(s)$ is non-empty, it is a lattice.*

This corollary is well-known in matching theory since Demange and Gale (1985) showed that the set of payoffs which ensures a stable allocation is a lattice whenever it is non-empty. 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, because the yogurt receives payoff $v_j = -\delta_j$. Conversely, the lower bound corresponds to the unanimously least preferred stable allocation for the consumers and the unanimously most preferred stable allocation for the yogurts ("yogurt-optimal")

The lattice structure of $\tilde{\sigma}^{-1}(s)$ implies that the set of systematic utilities rationalizing market shares s has a minimal and a maximal element; this structure is inherited from the equivalence of the NARUM and matching problems. Matching theory provides algorithms to compute these upper and lower extremal elements. Computing the minimal and maximal elements provides a very simple data-driven test of 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 "letting the data speak" as to whether a given model is point or partially-identified is an important virtue of the matching-style approach which we take in this paper.

3.3.2. *Existence.* 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 $\mathcal{U}_{\zeta j}(\delta)$ 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 $\Pr(\mathcal{U}_{\zeta j}(\delta^*) > -a) < \eta$.

(b) $\mathcal{U}_{\zeta j}(\delta)$ 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 $\Pr(\mathcal{U}_{\zeta j}(\delta^*) < b) < \eta$.

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 $s \in \mathcal{S}_0^{int}$.*

3.3.3. *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 nonvanishing density over $\mathbb{R}^J$.

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 nonvanishing density over $\mathbb{R}^J$ when $(\varepsilon_0 - \varepsilon_j)$ does.

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

Given uniqueness, we also consider convergence properties. 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 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 nonvanishing 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.*

4. Algorithms

The equivalence established in theorem 1 between matching and discrete choice models allows us to use matching algorithms in order to perform demand inversion in random utility models. In principle, 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 it would be very slow and inefficient, especially in the common situation where there are fewer products than consumers.

In the following, we present two types of algorithms. The algorithms of the first type, discussed in Section 4.1, are based on "equilibrium transport" a.k.a. matching with imperfectly transferable utility and therefore apply to all models – ARUMs or NARUMs. They solve for equilibrium in matching models, often by iteratively adjusting the payoffs of the potential partners, but as these problems are not optimization problems in general, they do not rely on optimization techniques. In contrast, the second class of methods, in Section 4.2, work specifically for additive random utility models and are based on "optimal transport" methods that themselves rely on optimization techniques – linear, combinatorial, or semidiscrete. Importantly, two of the following algorithms – namely, Market Shares Adjustment (paragraph 4.1.1) and Linear Programming (paragraph 4.2.1) – work regardless of whether the model is point-identified, returning the parameter itself in the point-identified case and upper and lower bounds on the parameter under partial identification. The other algorithms generally assume point identification of the utilities. However, we have augmented the auction algorithm (as discussed in paragraph 4.2.2) so that it provides lower and upper lattice bounds on the systematic utilities under partial identification.

The use of these algorithms in the empirical discrete-choice literature is new. However, as we will point out below, the IPFP (Iterative Proportional Fitting Procedure), in the case of ARUM models, turns out to be identical to procedures typically used to estimate the random coefficients logit demand model (cf. Berry, Levinsohn, and Pakes (1995), Nevo (2000)).

4.1. **For all models: Equilibrium transport algorithms.**

4.1.1. *Market shares adjusting algorithm (MSA).* The market shares adjusting (MSA) algorithm is an "accelerated" version of the deferred acceptance algorithm of Kelso and Crawford (1982) in which the utilities for all jars of yogurts of the same brand are adjusted simultaneously. While MSA does converge remarkably quickly in our experience, we have not yet formally proven convergence.

In our discrete choice setting, consumers bid in successive rounds for yogurts. Between rounds, the systematic utilities pertaining to the yogurts chosen by consumers are decreased by an adjustment factor, and bidding continues until a round is reached when the share of consumers choosing each yogurt is less or equal to its market share. Since this accelerated process can lead to "overshooting" (in which utilities move below their equilibrium values), the MSA algorithm involves successive *approximation phases* in which the adjustment factor is gradually made smaller and smaller before a final *convergence phase.* Two versions of the algorithm return either 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 bound will coincide; hence running both versions of this algorithm yields a data-driven assessment of whether the model is point- or partially-identified.

We approximate the distribution of consumer heterogeneity P into a sample of size N with uniform weights; there are therefore N consumers $i \in \{1, ..., N\}$ and we assume that there is a quantity $m_j \approx N s_j \in \mathbb{N}$ of yogurts of type j, where if needed, m_j has been rounded to an adjacent integer in such a way that $\sum_{j\in\mathcal{J}_0} m_j = N$. Throughout the algorithm we maintain the normalization $\delta_0 = 0$.

We present below the pseudocode for finding the lattice upper bound; Appendix A.5 contains the algorithm for the lattice lower bound. Define $\bar{\delta}_j = \sup_{i\in\{1,\ldots,N\}} \mathcal{U}_{ij}^{-1}\left(\mathcal{U}_{i0}\left(\delta_0\right)\right)$. Clearly, $\bar{\delta}$ is an upper bound for the stable payoffs and for the lattice upper bound.

Algorithm 1 (MSA upper bound)**.** *Begin main loop*

Take $\eta^{init} = 1$ *and* $\delta_j^{init} = \bar{\delta}_j$.

Repeat:

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

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

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

End main loop

Begin inner loop

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

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

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

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

If $\sum_i \pi_{i0} \geq m_0$, *then set* $\delta_j \leftarrow \delta_j + 2\eta$ *for all* $j \in \mathcal{J}$, *and* $\eta \leftarrow \eta/4$.

Else set $\delta_j \leftarrow \delta_j - \eta 1\left\{\sum_i \pi_{ij} > m_j\right\}$ *for all* $j \in \mathcal{J}$.

End While

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

End inner loop

4.1.2. *Iterative Proportional Fitting Procedure.* The second type is the Iterative Proportional Fitting Procedure (IPFP), which follow's McFadden (1989)'s proposal of solving a model with an independent Gumbel-distributed random noise term $T\zeta_{ij}$ which is added to the random utility $\mathcal{U}_{\varepsilon ij}\left(\delta_j\right)$ associated with alternative j. In general, we replace the problem of maximizing $\mathcal{U}_{\varepsilon ij}\left(\delta_j\right)$ over $j \in \mathcal{J}_0$ with the problem of maximizing $\mathcal{U}_{\varepsilon ij}\left(\delta_j\right) + T\zeta_{ij}$ over the same set of alternatives, where $T > 0$ is the temperature parameter, and ζ_{ij} is a Gumbel random variable which is independent of ε_i. Averaging the indirect utility of i over the noise yields a formula for the indirect utility $u_i = u_{\varepsilon_i}$ where the smooth-max has replaced the max, that is

$$u_i = T \log \sum_{j \in \mathcal{J}_0} \exp\left(\mathcal{U}_{\varepsilon_i j}\left(\delta_j\right)/T\right) \tag{4.1}$$

As a result, the demand map in the new model is given by $\sigma_j^T(\delta) = \mathbb{E}_P\left[\exp\left((\mathcal{U}_{\varepsilon j}\left(\delta_j\right) - u_\varepsilon)/T\right)\right]$. For a fixed T and for N draws of $\varepsilon_i \sim P$, this demand map can be simulated as:

$$\sigma_j^{T,N}(\delta) = \sum_{i=1}^{N} \frac{1}{N} \exp\left(\frac{\mathcal{U}_{\varepsilon_i j}\left(\delta_j\right) - u_i}{T}\right). \tag{4.2}$$

We then take $\sigma^{T,N}$ as an approximation of σ, and solve for δ such that $\sigma^{T,N}(\delta) = s$. Combining with (4.1), we see that (u_i, δ_j) solves the system

$$\left\{\begin{array}{l} \sum_{i=1}^{N} \frac{1}{N} \exp\left(\frac{\mathcal{U}_{\varepsilon_i j}(\delta_j) - u_i}{T}\right) = s_j \\ \sum_{j \in \mathcal{J}_0} \frac{1}{N} \exp\left(\frac{\mathcal{U}_{\varepsilon_i j}(\delta_j) - u_i}{T}\right) = \frac{1}{N} \end{array}\right. . \tag{4.3}$$

This system of equations is the same as the system of equations of equilibrium in imperfectly transferable utility (ITU) matching models described in Galichon, Kominers, and Weber (2014), who describe an algorithm to solve it. We have:

Algorithm 2 (General IPFP)**.** *Start with an initial guess of the vector* $\left(u_i^0\right)$*.*

Step $k \geq 0$*:*

For each $j \in \mathcal{J}_0$*, define* δ_j^{k+1} *as the solution in* δ_j *of the first set of equations defined by (4.3), taking the vector* $\left(u_i^k\right)$ *as given.*

For each $i \in \{1, ..., N\}$*, define* u_i^{k+1} *as the solution in* u_i *of the second set of equations, taking* $\left(\delta_j^{k+1}\right)$ *as given, namely* u_i^{k+1} *is defined by (4.1).*

Repeat until δ^k *and* δ^{k+1} *are close enough.*

When T is small, the method is accurate but may experience numerical difficulties as the arguments of the exponentials may become too large. To remedy this problem, one may use the log-sum-exp reparameterization from machine learning; see e.g. Peyré and Cuturi (2017). The idea is to replace the expression $\log \sum_{k \in K} e^{a_k}$, which may blow up if some of the a_k are large positive numbers, by the expression $\log \sum_{k \in K} e^{a_k - \max_{k' \in K}\{a_{k'}\}} + \max_{k \in K}\{a_k\}$, which cannot blow up because the arguments of the exponentials are nonpositive. This leads to:

Algorithm 2′ (IPFP log-sum-exp form)**.** *Start with an initial guess of u_i^0 and iterate the following over k until δ^{k+1} is close enough to δ^k:*

(a) For each $j \in \mathcal{J}_0$, define δ_j^{k+1} the root of $\varphi_j(\delta) = 0$, where

$$\varphi_j(\delta) = \begin{cases} T \log \sum_{i=1}^{N} \exp\left(\frac{\mathcal{U}_{\varepsilon_i j}(\delta_j) - u_i^k - \max_i \left\{\mathcal{U}_{\varepsilon_i j}(\delta_j) - u_i^k\right\}}{T}\right) \\ + \max_i \left\{\mathcal{U}_{\varepsilon_i j}(\delta_j) - u_i^k\right\} - T \log(N s_j) \end{cases} \tag{4.4}$$

(b) For each $i \in \{1, ..., N\}$, define u_i^{k+1} by

$$u_i = T \log \sum_{j \in \mathcal{J}_0} \exp\left(\frac{\mathcal{U}_{\varepsilon_i j}(\delta_j) - \max_{j' \in \mathcal{J}_0} \mathcal{U}_{\varepsilon_i j'}(\delta_{j'})}{T}\right) + \max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon_i j}(\delta_j). \tag{4.5}$$

While the updating formulas in algorithms 2 and 2′ are mathematically equivalent, algorithm 2′ will behave well for small values of T, while algorithm 2 will fail for small values of T. At the same time, the robust version of the algorithm (namely algorithm 2′) is much slower because it involves taking the maximum, taking logarithms and exponentials. This will be apparent below in example 5.1.

4.2. **For ARUM models only: Optimal transport algorithms.** The two algorithms described thus far handle random utility models in great generality, accommodating both additive and non-additive models. For additive models, additional optimization approaches for optimal transport models are available which can be much faster. We describe three in turn.

4.2.1. *Generic linear programming.* For ARUM models, the solution can be obtained from the linear program (3.3). Chiong, Galichon, and Shum (2016) propose a procedure based on a finite-dimensional discretization of this problem. Specifically, they propose drawing a random sample $\left\{\varepsilon^1, ..., \varepsilon^N\right\}$ from the distribution P and using a generic linear programming solver. The estimation is therefore carried out on the sampled version of the linear program

(3.3), which is

$$\inf_{u_i,\delta_j} \quad \sum_{i=1}^{N} \frac{1}{N} u_i - \sum_{j \in \mathcal{J}_0} s_j \delta_j \tag{4.6}$$

$$s.t. \quad u_i - \delta_j \geq \varepsilon_j^i \tag{4.7}$$

Linear programming software such as Gurobi Optimization (2016) are well-suited for solving this problem.[12] Note that these algorithms may return both the lower bound and the upper bound in models which are not point-identified.

4.2.2. *Auction algorithms.* Auction-type algorithms à la Bertsekas (1992) provide an alternative approach to linear programming methods. 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. We describe here the generalized version of the algorithm suited for asymmetric problems with similar objects described in Bertsekas and Castanon (1989). We consider similar objects in the sense that each consumer buys a different jar of yogurt, we denote by κ, which belongs to classes or brands j and $j(\kappa) = j$. To keep the analogy between the algorithm and an auction, we define the prices p_κ as negative systematic utilities: $p_\kappa = -\delta_\kappa$. The algorithm contains a bidding and of an assignment phase and it stops when no one is unassigned.

Algorithm 3 (Auction with Similar Objects)**.** *Start with an empty assignment and a given vector of prices p_κ and set a scale parameter $\eta > 0$.*

Bidding phase

(a) For each i that is unassigned, find the best jar $\kappa^\star$ which maximizes the utility:

$$\mathcal{U}_{\varepsilon_i \kappa^\star}(-p_{\kappa^\star}) = \max_{\kappa} \mathcal{U}_{\varepsilon_i \kappa}(-p_\kappa) \tag{4.8}$$

[12] For these packages, one must provide the software with the constraint matrix. Using the method described in Galichon (2016, chap. 3), we vectorize the constraints $u_i - \delta_j \geq \varepsilon_j^i$ in the column-major order, by stacking them varying i first, then j. Denote $z = (u, \delta)$. Letting $A = \begin{pmatrix} 1_N \otimes I_{|\mathcal{J}_0|} & I_N \otimes 1_{|\mathcal{J}_0|} \end{pmatrix}$, $c = (1_N^\intercal / N, -s)$ and $d = vec(\varepsilon)$, the constraints (4.7) are expressed as $Az \geq d$, and problem (4.6) becomes the maximization of $z^\intercal c$ over $z \in \mathbb{R}^{N+|\mathcal{J}_0|}$ subject to $Az \geq d$, a form we can pass directly to typical solvers.

and define the value of the second best jar which does not belong to the same brand

$$w_i = \max_{j(\kappa) \neq j(\kappa^\star)} \mathcal{U}_{\varepsilon_i \kappa}(-p_\kappa) \tag{4.9}$$

(b) Compute the bid:

$$b_{i\kappa^\star} = p_{\kappa^\star} + \mathcal{U}_{\varepsilon_i \kappa^\star}(-p_{\kappa^\star}) - w_i + \eta \tag{4.10}$$

Assignment phase

Jar κ is attributed to its highest bidder $i^\star$ breaking potentially an assignment and its price is raised:

$$p_\kappa := b_{i^\star \kappa} \tag{4.11}$$

Final step

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

$$\delta_j = - \min_{\kappa \in j(\kappa)} p_\kappa \tag{4.12}$$

The algorithm gets arbitrarily closer to the true solution as η is decreased. The performance of the algorithm is considerably improved by applying it several times, starting with a large value of η and gradually decreasing it to achieve the desired degree of accuracy.

In the numerical experiments, we use a variant of this algorithm called the General Auction developed and implemented by Walsh and Dieci (2017). The implementation requires careful tuning; in addition, we augment the algorithms by solving a residual network problem using the primal solution, which yields two improvements: (i) detecting a non-optimal solution to the primal problem (indicated by the presence of a negative cycle in the network); and (ii) providing lower and upper lattice bounds on the dual systematic utilities in cases of partial identification.

4.2.3. *Semi-discrete transport algorithms.* The final algorithm we study is specialized for solving the particular case of the pure characteristics demand model (discussed as Example 2.1 above), an important model which use in empirical work has been limited due to its computational difficulties. In this section we discuss an exact method of solving this model in the case 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), and only recently spectacularly developped by Kitagawa, Mérigot, and Thibert (2016) and Lévy (2015). The idea, exposited in chapter 5 of Galichon (2016), is that the optimal transport problem (3.3) can be reformulated as a finite-dimensional unconstrained convex optimization problem

$$\inf_{\delta\in\mathbb{R}^{\mathcal{J}_0}} F(\delta)\,, \text{ where } F(\delta) = \mathbb{E}_P\left[\max_{j\in\mathcal{J}_0}\{\delta_j + \varepsilon_j\}\right] - \sum_{j\in\mathcal{J}_0}\delta_j s_j. \tag{4.13}$$

Semi-discrete algorithms consist of a gradient or a Newton descent or 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.

4.3. **Implementation.** We have developed several R packages that provide fast and efficient implementations of the algorithms described in this paper. The packages 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.

First, we provide an R-based interface to Geogram by Lévy (2018). Geogram is a C++ library of geometric algorithms, with fast implementations of two- and three-dimensional semi-discrete optimal transport methods, which we use to estimate the pure characteristics model. The package is open-source and is available on GitHub as Lévy, Bonnet, Galichon, O'Hara, and Shum (2018).

Second, Bonnet, Galichon, O'Hara, and Shum (2018a) collects several auction and assignment algorithms into an R package, utilizing C++ code provided by Walsh and Dieci (2017). Finally, parallelized implementations of the MSA and IPFP algorithms are contained in a separate package Bonnet, Galichon, O'Hara, and Shum (2018b). Installation instructions can be found at the links provided in the references section.

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 in which the utilities are partially identified.

5.1. **Example 1: the pure characteristics model.** We recall from section 2 that a consumer i derives its utility from a product specific term (δ_j) and from individual preferences shocks (ν_i) related to the characteristics of the good (x_j):

$$u_{\varepsilon_i j} = \delta_j + \nu_i^{\intercal} x_j$$

In order to compare the performance of our algorithms, we run simulations with 5, 100 and $1,000$ brands of yogurt. The characteristics of the goods x_j have three dimensions and are drawn from standard uniform distributions. Taste shocks ν are also distributed as standard uniform. For all but the semi-discrete algorithm, we have to draw taste shocks to perform the estimation, and we used $10,000$ draws.

The pure characteristics model has not been applied much in empirical work due to its technical and computational challenges. Berry and Pakes (2007) suggest, as an approximation, to compute it using the usual BLP contraction mapping for random coefficient logit models, and so we also consider this approach as a benchmark in this example. It turns out (as we show in appendix A.6) the BLP contraction mapping is equivalent to the IPFP for ARUM models; for numerical accuracy we set the temperature parameter to a small value ($T = 0.01$) and therefore use the log-sum-exp version of the IPFP (Algorithm 2').

In table 1, we compare the efficiency of our algorithms with 5, 100 and $1,000$ brands. As the numerical accuracy is very comparable across all the algorithms, we focus on the computational time. The benchmark IPFP performs poorest, requiring the most computational time, and is unable to complete the 1000 brand model. The semi-discrete algorithm is the fastest and the General Auction comes second. However, currently, the semi-discrete algorithm can only solve problem with 3 or less characteristics. MSA's performance is also slow, but the main advantage of this algorithm is that it can be used in non-ARUM settings,

	IPFP	MSA	Lin.Prog.	Auction	Semi-discrete
5 brands	4.29	0.36	0.30	0.11	0.01
100 brands	962.37	846.98	8.81	0.75	0.11
1,000 brands	NA	NA	505.87	7.79	1.29

Note: The table displays the performance of the different algorithms in the context of the pure characteristics model with 5, 100 and $1,000$ brands of yogurt and $10,000$ draws of taste shocks. The algorithms considered are the IPFP in the log-sum-exp form (as a benchmark), MSA, linear programming, the General Auction, and the semi-discrete algorithm.

TABLE 1. Average computational Time (secs.) for pure characteristics models

and also for partially identified models. These virtues of the MSA are highlighted in the next example.

5.2. **Example 2: mixture vertical differentiation model.** As we remarked earlier, an important benefit of our approach is the ability to handle models which may not be point identified, as the algorithms we have presented above will yield estimates of the upper and lower bound values of the δ_y's even in the case when the model is not point identified. To see how this works, we consider a "mixture vertical differentiation" example; this is a model with continuous shock distribution which is not point identified.

There are three goods $y = 1, 2, 3$, and the unknown parameters are the quality of each good, given by $\delta_1, \delta_2, \delta_3$ Assume $\delta_1 = 0 < \min\{\delta_2, \delta_3\}$. 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:

$$s_y = 0.5(s_y^1 + s_y^2), \quad y = 1, 2, 3. \tag{5.1}$$

Consider goods y and y' with $\delta_y \geq \delta_{y'}$. Consumer i prefers to buy good y over y' at store j if and only if[13]

$$\mathcal{U}_{\varepsilon_i,y}(\delta_y) \geq \mathcal{U}_{\varepsilon_i,y'}(\delta_{y'}) \Leftrightarrow \theta \geq \hat{\theta}_j(y,y') \equiv \min\left\{1, \max\left\{0, \frac{p_y^j - p_{y'}^j}{\delta_y - \delta_{y'}}\right\}\right\} \tag{5.2}$$

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 the normalization $\delta_1 = 0$, the set of quality parameters (δ_2, δ_3) that rationalize these market shares is:

$$C = \{(\delta_2, \delta_3) : \delta_2 = 2, \delta_3 \in [1,3]\} \tag{5.3}$$

The set of quality parameters which rationalize the observed market shares is thus a lattice, with join $(0, 2, 1)$ and meet $(0, 2, 3)$.

Computational results from this example, using the market shares adjustment with scaling algorithm (which can be adapted to partially identified models), are given in Table 2. The algorithm performs as expected, and running the two versions of the MSA algorithm (described in Section 4.1.1 and Appendix A.5) correctly recovers the upper and lower bounds.

	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.

TABLE 2. Mixture vertical differentiation: 3 goods

[13]Note that, according to (5.2), 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.

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 3. Since, these market shares were chosen arbitrarily, we do not *a priori* whether the utilities vectors δ are point- or partially-identified.[14] Hence we will compute both the consumer- and yogurt-optimal versions of this algorithm, to determine this, and also to recover the relevant (range of) utilities. In Table 4, we report the estimated utilities.

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

TABLE 3. Mixture vertical differentiation, 5 stores and 8 goods

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 4 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.

6. 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

[14] 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 3 arbitrary, and use our approach to determine whether the utility parameters are point- or partially-identified.

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

TABLE 4. Lower bound and upper bound in the mixture vertical differentiation with 5 stores and 8 goods

imperfectly transferable utility to derive identification and estimation procedures for discrete choice models based on the non-additive random utility specification.

Although it is commonplace to distinguish in the microeconomic literature 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 indeed just as appropriate to consider a discrete choice problem as one in which consumers choose yogurts, as one in which (fancifully) "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[15] 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.

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

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.

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.

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

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

Berry, S., O. B. Linton, and A. Pakes (2004): "Limit theorems for estimating the parameters of differentiated product demand systems," *Review of Economic Studies*, 71(3), 613–654.

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

——— (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.

Chiappori, P., I. Komunjer, and D. Kristensen (2009): "On Nonparametric Identification and Estimation 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.

Choo, E., and A. Siow (2006): "Who marries whom and why," *Journal of Political Economy*, 114(1), 175–201.

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.

Dagsvik, J. K. (2000): "Aggregation in matching markets," *International Economic Review*, 41(1), 27–58.

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

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.

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 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., S. D. Kominers, and S. Weber (2014): "An empirical framework for matching with imperfectly transferable utility," *Available at SSRN 2535394*.

Galichon, A., and B. Salanié (2017): "Cupids invisible hand: Social surplus and identification in matching models," working paper, NYU.

Gurobi Optimization, I. (2016): "Gurobi Optimizer Reference Manual," .

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.

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.

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," `http://alice.loria.fr/index.php/software/4-library/75-geogram.html`.

Lévy, B., O. Bonnet, A. Galichon, K. O'Hara, and M. Shum (2018): "Rgeogram - R Wrapper to the Geogram Library," `https://github.com/TraME-Project/Rgeogram`.

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

McFadden, D. (1989): "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, pp. 995–1026.

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

Menzel, K. (2015): "Large Matching Markets as Two-Sided Demand Systems," *Econometrica*, 83(3), 897–941.

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.

Nöldeke, G., and L. Samuelson (2018): "The implementation duality," *Econometrica*, 86, 1283–1324.

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*.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and their Applications," *Econometrica*, 83, 315–334.

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.

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.

Roth, A. E., and M. A. O. Sotomayor (1992): *Two-sided matching: A study in game-theoretic modeling and analysis*, no. 18. Cambridge University Press.

Rust, J. (1987): "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica*, pp. 999–1033.

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.

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*.

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.[16] 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 $\Sigma(\delta)$.

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

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: Conversely, 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.[17] ∎

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.[18]

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

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

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

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

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 $\overline{\sigma}_j(\delta) = 1$. □

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

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

A.2. **Proof of theorems 1–4.**

A.2.1. *Proof of theorems 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,

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

In particular, for each open set B, one has

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

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

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

for all $\left(\varepsilon,j\right)$, with equality on the support of π. Hence, conditions (ii) and (iii) in definition 3 are met, and $\left(\pi,u,v\right)$ is an equilibrium outcome.

(b) From equilibrium matching to demand inversion: Let $\left(\pi,u,v\right)$ be an equilibrium matching in the sense of definition 3, where $f_{\varepsilon j}\left(x\right)=x$ and $g_{\varepsilon j}\left(y\right)=-\mathcal{U}_{\varepsilon_{i}j}\left(-y\right)$. Then letting $\delta=-v$, one has by condition (ii) that for any $\varepsilon\in\Omega$ and $j\in\mathcal{J}_{0}$,

$$u\left(\varepsilon\right)-\mathcal{U}_{\varepsilon j}\left(\delta_{j}\right)\geq 0$$

thus $u\left(\varepsilon\right)\geq\max_{j\in\mathcal{J}_{0}}\mathcal{U}_{\varepsilon j}\left(\delta_{j}\right)$. But by condition (iii), for $j\in Supp\left(\pi\left(.|\varepsilon\right)\right)$, one has $u\left(\varepsilon\right)=\mathcal{U}_{\varepsilon j}\left(\delta_{j}\right)$, thus

$$u\left(\varepsilon\right)=\max_{j\in\mathcal{J}_{0}}\mathcal{U}_{\varepsilon j}\left(\delta_{j}\right).$$

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

$$\sigma_{j}\left(\delta\right)=P\left(\varepsilon\in\Omega:J\left(\tilde{\varepsilon}\right)=\left\{j\right\}\right)=\Pr\left(\tilde{j}=j\right)=s_{j}.$$

Hence $s\in\tilde{\Sigma}\left(\delta\right)$, 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 $s\in\tilde{\Sigma}\left(\delta\right)$ and $s'\in\tilde{\Sigma}\left(\delta'\right)$, then*

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

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[19]. □

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.

$$\delta_j^\wedge = \min\left(\delta_j, \delta'_j\right) \text{ and } \delta_j^\vee = \max\left(\delta_j, \delta'_j\right),$$

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

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

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

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

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

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

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

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

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

(b) Proof of $s' \in \tilde{\Sigma}\left(\delta \vee \delta'\right)$: 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 $s' \in \tilde{\Sigma}\left(\delta \vee \delta'\right)$, it is sufficient to show that for any $B \subseteq \Omega$,

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

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

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

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

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

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

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

is supermodular in δ and has increasing differences in (δ, s). As a result of Topkis' theorem, the set-valued map $s \to \tilde{\Sigma}^{-1}(s)$ 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 $s \in \mathcal{S}_0^{int}$.*

Remark A.5. While in this paper we make use of theorem $3'$ to guarantee the existence of an identifying systematic utility vector δ under very weak restrictions, this result is a contribution to matching theory per se. Indeed, it implies the existence of a solution to the equilibrium transport problem, as introduced in Galichon (2016), definition 10.1.

The proof is based on several lemmas.

Assumption 4 implies that for all $\eta > 0$ and $\nu > 0$ there is δ^* s.t. $\delta > \delta^*$ implies $\Pr\left(|X_\delta - X_\delta^*| > \nu\right) < \eta$.

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

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

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:

Fact (a): $F_j^T(.)$ 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

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

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 $\delta_j^{T,k+1}\in(\delta_j^{T,k},\delta_j^*)$ 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,

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

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

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

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

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

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

and by the same result, one has

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

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 $\pi^0 \in \mathcal{M}(P, s)$.

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

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

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

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

Start by showing point (i). We have $0 \leq u^0(\varepsilon) - u^T(\varepsilon) \leq T_n \log \mathcal{J}$. 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

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

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 $n \geq \overline{n}$ implies $\max_{j\in\mathcal{J}} \left|\mathcal{U}_{\varepsilon J}\left(\delta_j^{T_n}\right) - \mathcal{U}_{\varepsilon J}\left(\delta_j^0\right)\right| \leq \nu/2$ for each $\varepsilon \in K$. Thus, for $n \geq n'$, one has

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

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

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

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

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

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 $u^0(\varepsilon) \geq \mathcal{U}_{\varepsilon j}\left(\delta_j^0\right)$ 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 $Z_j = \mathcal{U}_{\varepsilon j}^{-1}(\mathcal{U}_{\varepsilon 0}(\delta_0))$, and P_Z to be the probability distribution of Z. By proposition 1, this implies

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

which is equivalent to

$$P(Z_j > \delta_j) \leq s_0 \leq P(Z_j \geq \delta_j)$$

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

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

Now consider $\delta^{\min}$ and $\delta^{\max}$ the lattice bounds of $\tilde{\Sigma}^{-1}(s)$. Because of the previous remark, $P\left(Z_j \geq \delta_j^{\min}\right) = P\left(Z_j \geq \delta_j^{\max}\right)$. 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 $j \notin B$

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

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* $\max\left(\|\delta\|, \|\delta'\|\right) \leq \rho$*, one has*

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

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. ∎

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. ∎

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

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

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

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

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 should be related 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).

Remark A.6. (Equivalence of BLP contraction mapping and IPFP) In the random coefficient logit model, one has $\mathcal{U}_{\varepsilon ij}(\delta_j) = \delta_j + \nu_i^{\intercal} x_j + T\zeta_{ij}$, where $\nu \sim \mathbf{P}_\nu$ and (ζ_j) has an i.i.d. Gumbel distribution, and both these vectors are distributed independently. As a result, the system (4.3) becomes

$$\begin{cases} \sum_{i=1}^{N} \frac{1}{N} \exp\left(\frac{\delta_j - u_i + \nu_i^{\intercal} x_j}{T}\right) = s_j \\ \sum_{j\in\mathcal{J}_0} \frac{1}{N} \exp\left(\frac{\delta_j - u_i + \nu_i^{\intercal} x_j}{T}\right) = \frac{1}{N} \end{cases}. \tag{A.26}$$

For this case, the IPFP algorithm (algorithm 2) consists of applying the following updating rules

$$\delta_j^{k+1} = -T\log\left(\sum_{i=1}^{N} \frac{1}{N s_j} \exp\left(\frac{-u_i^k + \nu_i^{\intercal} x_j}{T}\right)\right),\ j \in \mathcal{J} \tag{A.27}$$

$$u_i^{k+1} = T\log\left(\sum_{j\in\mathcal{J}_0} \exp\left(\frac{\delta_j^{k+1} + \nu_i^{\intercal} x_j}{T}\right)\right),\ 1 \le i \le N \tag{A.28}$$

This algorithm can be viewed as coordinate descent over δ and u iteratively. Indeed, the optimality conditions with respect to δ_j holding u^k constant lead to the first set of updating formulas, while the optimality conditions with respect to u_i holding δ^{k+1} fixed lead to the second set of updating formulas.

Berry, Levinsohn, and Pakes (1995) (formula 6.8 p. 865) propose an algorithm which consists of the following updating formula (in the paper they normalize $T = 1$, but the extension is straightforward)

$$\delta_j^{k+1} = \delta_j^k + T\log s_j - \log \sum_{1\le i\le N} \frac{1}{N} \frac{\exp\left(\frac{\delta_j^k + \nu_i^{\intercal} x_j}{T}\right)}{\sum_{j'\in\mathcal{J}_0} \exp\left(\frac{\delta_{j'}^k + \nu_i^{\intercal} x_{j'}}{T}\right)}.$$

The IPFP algorithm in the additive case, which is defined by formulas (A.27) and (A.28), is equivalent to BLP's contraction mapping algorithm. Indeed, if we define at each step $u_i^k = T\log\left(\sum_{j\in\mathcal{J}_0} \exp\left(\frac{\delta_j^k + \nu_i^{\intercal} x_j}{T}\right)\right)$ as in (A.28), we retrieve exactly formula (A.27). Hence, algorithm 2 is an extension of BLP's contraction mapping algorithm to the nonadditive case.

A well-known problem with BLP's contraction mapping algorithm is its behavior under small T as exposed in Berry and Pakes (2007), as the exponentials may blow up if the arguments are large positive numbers. Fortunately, our IPFP algorithm with the log-sum-exp trick (algorithm $2'$) addresses this issue. We describe it in the additive case:

Algorithm $2''$ (BLP's algorithm–log-sum-exp form)**.** *Start with an initial guess of u_i^0 and iterate the following over k until δ^{k+1} is close enough to δ^k:*

$$\begin{cases} \bar{\delta}_j^{k+1} = \min_i \left\{ \frac{u_i^k - \nu_i^\intercal x_j + T \log s_j}{T} \right\} \\ \delta_j^{k+1} = \bar{\delta}_j^{k+1} - T \log \left(\sum_{i=1}^N \frac{1}{N} \exp \left(\frac{-u_i^k + \nu_i^\intercal x_j - T \log s_j + \bar{\delta}_j^{k+1}}{T} \right) \right) \\ \bar{u}_i^{k+1} = \max_{j \in \mathcal{J}_0} \frac{\delta_j^{k+1} + \nu_i^\intercal x_j}{T} \\ u_i^{k+1} = \bar{u}_i^{k+1} + T \log \left(\sum_{j \in \mathcal{J}_0} \exp \left(\frac{\delta_j^{k+1} + \nu_i^\intercal x_j - \bar{u}_i^{k+1}}{T} \right) \right) \end{cases}.$$

However, for the case where $T >> 0$ (as in BLP's original contraction mapping algorithm), the IPFP can be speeded up by putting it in the so-called Sinkhorn form,[20] which has the advantage that the updates formulas only involve linear algebra. Set $K_{ij} = \exp\left(\frac{\nu_i^\intercal x_j}{T}\right)$ and $a_i = N^{-1} \exp(-u_i/T)$, and $b_j = \exp(\delta_j/T)$. IPFP in the Sinkhorn form is computationally equivalent to the contraction mapping algorithm described in Nevo (2000) which solves for the exponent of the systematic utility δ in a BLP setting. In this case, the updating formulas become:

Algorithm $2'''$ (BLP's algorithm–Sinkhorn form)**.** *Start with an initial guess of $a_i^0 > 0$ and iterate the following over k until $T \log b^{k+1}$ is close enough to $T \log b^k$:*

$$\begin{cases} b_j^{k+1} = \frac{s_j}{(K^\intercal a^k)_j} \\ a_i^{k+1} = \frac{1/N}{(K b^{k+1})_i} \end{cases}.$$

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, it would not be efficient since the problem is generally asymmetric: there are few brands of yogurts and a lot of 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

[20] Peyré and Cuturi (2017), chapter 4

to values of δ which are below the lower bound $\underline{\delta}$. In the "second loop", we iterate up from this point up to the lower bound.

Algorithm 4 (MSA lower bound)**.** *Take* $\eta^{init} = 1$, $\delta_j^{init} = \delta_j^{ub}$ *and* $block_j = 0$ *for all* $j \in \mathcal{J}$.

Begin first loop

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 π_{ij} *equal to one if* $j \in \arg\max_j \mathcal{U}_{\varepsilon j}(\delta_j)$, *zero otherwise (breaking ties arbitrarily if needed).*

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

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

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

Else $block_j \leftarrow block_j$

End While

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

End first loop

Begin main second loop

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

Repeat:

Call the inner 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* $j \in \mathcal{J}$.

End main second loop

Begin inner second loop

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

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

Set π_{i0} *equal to one if* $0 \in \arg\max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j)$, *zero otherwise (breaking ties arbitrarily if needed).*

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

Set π_{ij} *equal to one if* $j \in \arg\max_j \mathcal{U}_{\varepsilon j}(\delta_j)$, *zero otherwise (breaking ties arbitrarily if needed).*

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

Set π_{i0} *equal to one if* $0 \in \arg\max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j)$, *zero otherwise (breaking ties arbitrarily if needed).*

End While

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

End inner second loop