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

ODRAN BONNET[♭], ALFRED GALICHON[†], AND MATT SHUM[§]

Abstract. In this paper we describe an equivalence between random utility discrete-choice models and two-sided matching models with imperfectly transferable utility, and we exploit its consequences. Based on it, we suggest new approaches for estimation and identification of non-additive random utility models (NARUM), in which the utility shocks do not affect decision-makers' utilities in an additive manner. The estimation algorithms and procedures we describe are inspired by those in the matching literature. A noteworthy feature of our algorithms is that they yield the point estimate when the model is point identified, and yield the upper and lower bounds on the parameters under partial identification.

Keywords: non-additive random utility models, two-sided matching, deferred acceptance, partial identification

JEL Classification: C51, C60

PRELIMINARY AND INCOMPLETE VERSION

Date: March 5, 2017 (First draft: 9/2015). Galichon gratefully acknowledges funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreements no 313699. We thank Jeremy Fox and participants in seminars at Johns Hopkins, UNC and the 2015 Banff Applied Microeconomics Conference for helpful comments. Alejandro Robinson-Cortes provided excellent research assistance.

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

[†]New York University; alfred.galichon@gmail.com

[§]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 $j \in \mathcal{J}$ where alternative j yields utility U_{ij} to agent i. 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 a demand shifter δ_j, or mean utility level associated with alternative j, which is identical across all agents, and a realization ε_{ij} of agent i's idiosyncratic preference, or "utility shocks" vis-à-vis alternative j. The agent's program is then

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

Note that particular instances of these models are additive random utility models (ARUM), such as the Logit or Probit models, where $\mathcal{U}_{\varepsilon_i j}(\delta_j) = \delta_j + \varepsilon_{ij}$; however, our setting does not require to impose this restriction. Typically, frameworks considering time or risk preferences are naturally non additive[1].

The *demand map* $\sigma(\cdot)$ is defined as the probability that alternative j dominates all the other ones, given the vector of systematic utilities $(\delta_j)_{j \in \mathcal{J}}$. Hence,

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

The question we address in this paper is one of *demand inversion*: given observed market shares $(s_j)_{j \in \mathcal{J}}$, how can one characterize the full set of utility vectors $(\delta_j)_{j \in \mathcal{J}}$ such that $s = \sigma(\delta)$? We need to address a *partial identification* issue here, as the identified set of vectors δ that are solution to the demand inversion problem may not necessarily be restricted to a single point. Also the characterization we are looking for should be *constructive*, as we are looking for efficient computational techniques to construct this set.

To answer this question, we establish an *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

[1] Apesteguia and Ballester (2014)

always be interpreted as a matching market where consumers and alternatives are interpreted as workers and firms; and that the *identification problem* of demand inversion can be reformulated as the *equilibrium problem* of determining Walrasian 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 identification results arising from the equivalence between discrete-choice models and two-sided matching are flexible and powerful, and allow us to move beyond ARUMs to more general class of non-additive random utility models (NARUMs), in which $\mathcal{U}_{\varepsilon_{ij}}(\delta_j)$ is no longer necessarily the sum of the systematic utility term δ_j and the idiosyncratic utility shock ε_{ij}.

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

(1) **Characterization of the structure of the identified set δ.** It follows from the equivalence principle that the identified set is a *lattice*. Much of the existing literature on discrete-choice models proceeds by defining 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 lattice greatest element coincides with the lattice smallest element, then the set is point-identified. Additionnaly, we introduce different assumptions which ensures connectedness, existence, point-identification and stability.

(2) **Computation of the identified set using matching algorithms**. The other benefit of the equivalence principle and reformulation as a matching is that the greatest and smallest elements of the identified set can be computed by matching algorithms, among which the celebrated *deferred acceptance algorithm* with salaries of Crawford and Knoer (1981). The algorithms we propose are valid no matter if

the underlying model is point- or partially-identified, and they will recover either point- or set-identified values of δ depending on what the case may be.[2]

Besides these results which contribute to the literature on discrete choice models, our paper makes a contribution to the theory of two-sided matching with imperfectly transferable utility. In particular, the results of section 4 on the isotonicity of the set of matching payoffs δ with respect to the marginal distribution s, the existence of an equilibrium, as well as the stability of the equilibrium, are novel results in this context.

1.1. **Existing literature.** The simplest setting for demand inversion is the logit model, which is an additive random utility model with utility shocks ε_{ij} 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 renowned "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 substitutions patterns. This has led the literature to investigate more general specifications, either within ARUMs or within NARUMs. In this more general setting, an explicit inversion formula such as (1.3) does not exist.

Following Berry (1994) and Berry, Levinsohn, and Pakes (1995) many authors have performed nonparametric demand inversion when the distribution of utility shocks is known, at least up to a parameter. There are several reasons for wishing to nonparametrically recover the full vector of utilities (δ_j) from the observed market shares (s_j). A number of authors have demonstrated the strength of this approach in dynamic discrete choice problems à la Rust (1987) and have developped 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 in order to deal with endogeneity problems (Berry, Levinsohn, and Pakes 1995, Berry and Haile 2014).

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

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 making 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 based on Linear Programming for computing the utility vector based on the market shares.

It is thus well known in the literature that in the additive case and when the utility shocks are continuously distributed and have full support, 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) remark, "Unlike the inversion results for the parametric examples, 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 studied 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 matching algorithms such as the one developped in Crawford and Knoer (1981) and Kelso Jr and Crawford (1982) which are generalizations of Gale and Shapley's deferred acceptance algorithm. The underlying mathematical structure at the core of the duality between these problems is the notion of a "Galois connection", recently introduced in economics for principal-agent problems by Noldeke and Samuelson (2015).

1.2. **Organization.** Section 2 presents the general NARUM framework which is the focus in this paper, and provides examples and comparisons with the more restrictive ARUM framework. Section 3 introduces the equivalence of NARUM with two-sided matching problems and discusses identification and estimation in this context. Section 4 reveals the lattice structure of the identfied set of utility vectors which arises from the inverse isotonicity of the demand map $\sigma()$. Subsequently, in Section 5, we tackle the issues of existence, uniqueness and stability. In Section 6, we present three computational algorithms, and we develop two numerical investigations of the algorithms.

2. The framework

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. Indeed, an explicit treatment of this alternative 0 is important to our approach (see below section 4.1 about normalization). The agent's program in a Nonadditive Random Utility Model (NARUM) is thus

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

where $u\left(\varepsilon_i\right)$ is the indirect utility of an agent with shock ε_i. We assume that the functions $\mathcal{U}_{\varepsilon j}\left(\cdot\right)$, as well as the distribution of ε_i for each agent i, are known to the researcher. Thus we focus throughout on parametric NARUM models. The utility an agent i derives from alternative j depends on the demand shifter 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 demand shifter 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 $\varepsilon \mapsto \left(\mathcal{U}_{\varepsilon j}\left(\delta_j\right)\right)_{j \in \mathcal{J}_0}$ is measurable, and

(b) the map $\delta_j \mapsto \mathcal{U}_{\varepsilon j}\left(\delta_j\right)$ 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 demand shifters δ, there is zero probability that an agent is indifferent between between two alternatives. For large parts of this paper (namely, parts of section 3 and 4), we will *not* invoke this assumption.

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',*

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

Assumptions 1 and 2 ensure that the market share $\sigma_j\left(\delta\right)$ of alternative $j \in \mathcal{J}_0$ predicted by utility vector δ is well defined and the market share of alternative j can be defined as the fraction of consumers who prefer weakly *or* strictly that alternative to any other one:

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

In the following, we also consider situations where Assumption 2 is violated, implying that agents may be indifferent between available alternatives. Violations of Assumption 2 involve partial identification of the NARUM model, which is an important topic that we cover in this paper.

Next, we consider several examples of NARUM.

Example 2.1 (ARUM)**.** Obviously, the classic additive random utility model (ARUM) is a NARUM. In the ARUM, $\Omega = \mathbb{R}^{\mathcal{J}_0}$, and P is an absolutely continuous probability distribution on $\mathbb{R}^{\mathcal{J}_0}$, and

$$\mathcal{U}_{\varepsilon j}\left(\delta_j\right) = \delta_j + \varepsilon_j.$$

In particular, 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 get the Logit model, where $\sigma_j\left(\delta\right) = \exp\left(\delta_j\right) / \left(\sum_{j' \in \mathcal{J}_0} \delta_{j'}\right)$.

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. Assume that ε_i is the relative risk aversion parameter (say utility is

CRRA) and does not depend on the alternative j, and that the price of good j is p_j. Making the choice 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

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

where the expectation is taken over η_j holding ε_i constant. Apesteguia and Ballester (2014) have highlighted the flaws of the classical techniques of estimation to estimate these kind of models.

Example 2.3 (Vertical differentiation model)**.** We consider an example of the classic vertical differentiation demand framework.[3] Assume that household i obtains utility from brand j equal to

$$\epsilon_i * \delta_j - p_j, \quad \forall j.$$

Here δ_j is interpreted as the quality of brand j, while the nonlinear random utility shock ϵ_i measures household i's willingness-to-pay for quality. Below, in section 6, we will consider a numerical example based on this framework.

Example 2.4 (Indirect utility from consumption)**.** Assume that utility from consumption basket x is $V(x) + \delta_0$ 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

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

which is a binomial NARUM.

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

3. The Demand Inversion Problem is a Matching Problem

In this section, we formalize our central equivalence result between the demand inversion problem and the determination of a stable outcome in a certain two-sided matching problem. To build intuition, we will start in section 3.1 by working under assumptions 1 and 2, the latter assumption guaranteeing the existence of a demand function σ defined in (2.2). In section 3.2, we shall see that the result extends to the more general case where assumption 2 is dropped, in which case, indifference is possible, the demand function σ is no longer defined, but should be replaced by a set-valued analog.

3.1. The equivalence result without indifferences.

In this section, we shall maintain assumptions 1 and 2. The demand inversion problem consists in, given a vector of observed maket 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 1 (Identified utility set)**.** Given a demand map σ defined as in (2.2) 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

$$\sigma^{-1}(s) = \left\{ \delta \in \mathbb{R}^{\mathcal{J}_0} : \sigma(\delta) = s \right\}. \tag{3.1}$$

The discrete choice problem is traditionally considered a one-sided problem, as it is evident that consumers choose yoghurts, but yoghurts are inanimate and do not choose consumers. However, we will now describe an equivalence with a two-sided problem: a "marriage problem" between consumers and yoghurts, where both sides of the market must assent to be matched. 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.

Let us describe this equivalent matching game between consumers and yoghurts.[4] Let $\mathcal{M}(P,s)$ be the set of probability distributions on $\Omega \times \mathcal{J}_0$ with marginal distributions P and s. 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 yoghurt j. Symetrically, let $g_{\varepsilon j}(v)$ be the transfer needed by a yoghurt 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.

Definition 2 (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) there is pairwise feasibility: if $(\varepsilon, j) \in Supp(\pi)$, then $f_{\varepsilon j}(u_\varepsilon) + g_{\varepsilon j}(v_j) = 0$.

Let us briefly comment on these three conditions. 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 yoghurt of type j. As a result, $\pi(j|\varepsilon)$ is the conditional choice probability that an individual with utility shock ε chooses yoghurt j. The second condition implies that if an individual with utility shock ε and a yoghurt 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 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 : $f_{\varepsilon j}(u) + g_{\varepsilon j}(v) = 0$.

[4]This model was introduced by Demange and Gale (1985) to extend the model of Shapley and Shubik (1971) beyond the transferable utility setting; Chapter 9 in Roth and Sotomayor (1992) also exposits this model. 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 unchanged.

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. *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 $\delta \in \mathbb{R}^{\mathcal{J}_0}$. Then, the two following statements are equivalent:*

(i) δ belongs to the identified utility set $\sigma^{-1}(s) = \left\{\delta \in \mathbb{R}^{\mathcal{J}_0} : \sigma(\delta) = s\right\}$ associated with the market shares s in the sense of definition 1 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 2 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 the NARUM and a problem of matching with partially transferable utility in the manner of Crawford and Knoer (1981),Kelso Jr and Crawford (1982), Hatfield and Milgrom (2005), 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 theorem is that in a matching equilibrium, the transfers are adjusted so that everyone is happy with their own choices: a consumer ε will be matched with the yoghurt j which maximizes her utility given v_j. Hence in the matching model, 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 it 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 feasibility 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 no blocking pair condition is satisfied : if $(\varepsilon, j) \in Supp(\pi)$ then $u_\varepsilon = \max\limits_{j'\in\mathcal{J}_0} \mathcal{U}_{\varepsilon j'}\left(\delta_{j'}\right) = \mathcal{U}_{\varepsilon j}(\delta_j)$. However, it is not immediate to show that the condition $\sigma(\delta) = s$ implies the existence of a $\pi \in \mathcal{M}(P, s)$

such that $(\pi, u, -\delta)$ is a matching equilibrium. This part relies on a probabilistic result called Strassen's (1965) theorem, which is essentially an infinite-dimensional extension of Hall's marriage lemma.

Moreover, note that the joint distribution $\pi \in \mathcal{M}(P, s)$ is interpreted in the matching theory as the probability of finding a pair (ε, j) matched at equilibrium, and is interpreted in the discrete choice setting as the probability of drawing an agent of type ε and that the choice of this agent is j. In other terms, the conditional probability $\pi(j|\varepsilon)$ is the conditional choice probability of choosing alternative j conditional on drawing type ε. (Under Assumption 2, this probability will be either 0 or 1.)

3.2. **The equivalence result with indifferences.** In this subsection, we investigate the case when we drop assumption 2, i.e. indifference between two alternatives may occur with positive probability. In this case, the random set of alternatives preferred by the agent no longer almost surely contains a single element. This is for instance the case in an additive random utility model with discrete heterogeneity; indifferences can also arise in vertical differentiation models (see Section 6.2.2). 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 3 (Demand correspondence)**.** The demand correspondence at δ, denoted $\Sigma(\delta)$, 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.

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

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

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

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

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.[5]

Remark 3.1. A necessary condition for the second statement of proposition 1 to hold is

$$s_j \leq P\left(U_{\varepsilon j}\left(\delta_j\right) \geq \max_{j' \in \mathcal{J}_0} U_{\varepsilon j'}\left(\delta_{j'}\right)\right), \quad \text{for all } j \in \mathcal{J}_0 \tag{3.3}$$

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 (3.3) 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 3.2. In general $\Sigma(\delta) \subseteq \left[\underline{\sigma}_j(\delta), \overline{\sigma}_j(\delta)\right]$, where we define

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

For instance, if $\mathcal{J} = \{j_1, j_2\}$ and $\mathcal{U}_{\varepsilon j}\left(\delta_{j_1}\right) = \mathcal{U}_{\varepsilon j}\left(\delta_{j_2}\right) = \mathcal{U}_{\varepsilon j}\left(\delta_{j_0}\right)$ 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 3.3. 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. □

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

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. The following result is then a direct extension of theorem 1 when assumption 2 is dropped; the only difference is that the inverse demand map $\sigma^{-1}(s)$ has been replaced by the inverse demand correspondence $\Sigma^{-1}(s)$.

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 1 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 2 in the matching problem, where

$$f_{\varepsilon j}(u) = u \text{ and } g_{\varepsilon j}(-\delta) = -\mathcal{U}_{\varepsilon j}(\delta). \tag{3.4}$$

4. Structure of the identified set

In this section, we derive several structure features of the identified set of utilities, which are practically important for estimation of these models. In particular, we derive that, for a given set of market shares, the identified set of utilities is a *lattice*; this implies, roughly speaking, that the set of identified utilities has a "largest" (resp. "smallest") vector 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;[6] 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.

[6]Among many others, see Ciliberto and Tamer (2009) and Pakes, Porter, Ho, and Ishii (2011).

These results hinge upon an important property of the map σ (and by extension, of the set-valued map Σ), namely the isotonicity of the inverse correspondence Σ^{-1}.[7] In the present setting, both $\Sigma(\delta)$ and $\Sigma^{-1}(s)$ can potentially be set-valued. Indeed, under assumption 2 the demand correspondance Σ is point-valued, but additional assumptions are needed to ensure bijectivity and therefore uniqueness (this question is investigated in section 5).

In our general setting, we need to extend the notion of monotonicity in order to make sense of the fact that Σ^{-1} should be isotone as a set-valued map. The right concept is Veinott's (2005) strong set order, which is discussed below. Theorem 2 establishes that, under only assumption 1, Σ^{-1} is isotone in the strong set order.

This result has several important practical consequences for the underlying identification problem. First, that the set of utilities rationalizing a vector of market shares s has a *lattice structure*, as will be expressed in theorem 3. The lattice structure of the identified set of utility vectors is important as it implies that this set has a greater and a smaller element. Also, the inverse isotonicity result allows to provide bounds on the vector of identified market share, as shown in proposition 3.

Before we discuss this, we first discuss the issue of the normalization of the utility associated to alternative 0.

4.1. **Normalization.** Any discrete-choice model requires some normalization, because the choice probabilities result from the comparison of the *relative* utility payoffs from each alternatives. In the sequel, we shall normalize the systematic utility associated to the default alternative, that is, we shall impose thereafter

$$\delta_0 = 0, \tag{4.1}$$

and we shall let $\mathcal{D}_0$ be the set of $(\delta_j)_{j\in\mathcal{J}_0}$ such that $\delta_0 = 0$, and $\mathcal{S}_0$ be the set of $(s_j)_{j\in\mathcal{J}_0}$ such that $\sum_{j\in\mathcal{J}_0} s_j = 1$. Consequently, we introduce $\tilde{\Sigma}$ as the restriction of Σ to $\mathcal{D}_0$, and hence $\tilde{\Sigma}^{-1}(s)$ is the set of market shares δ that rationalize choice probabilities s while satisfying

[7]This result is related to the result of Berry, Gandhi, and Haile (2013) who show, under conditions guaranteeing that both Σ and Σ^{-1} are point-valued, that inverse demand function is indeed isotone on its domain, hence the systemic utility vector δ is unique.

restriction (4.1), that is

$$\tilde{\Sigma}^{-1}(s) = \left\{\delta \in \mathbb{R}^{\mathcal{J}_0} : s \in \Sigma(\delta) \text{ and } \delta_0 = 0\right\}.$$

Remark 4.1. In the special case of an ARUM, imposing normalization (4.1) is innocuous in the sense that it is straightforward to recover $\Sigma^{-1}(s)$, the unnormalized set of $\delta \in \mathbb{R}^{\mathcal{J}_0}$ that rationalize the market shares s, on the basis of $\tilde{\Sigma}^{-1}(s)$, the normalized corresponding set. Indeed, $\delta \in \Sigma^{-1}(s)$ if and only if there is $c \in \mathbb{R}$ and $\tilde{\delta} \in \tilde{\Sigma}^{-1}(s)$ such that $\delta = \tilde{\delta} + c$. However, in the case of a NARUM, such a simple relationship is lost, so the normalization (4.1) is no longer innocuous: indeed, there exists no simple analytic relation between $\delta \in \Sigma^{-1}(s)$ such that $\delta_0 = 0$ and $\delta \in \Sigma^{-1}(s)$ such that $\delta_0 = c$. Therefore, the researcher should justify carefully his choice of normalization. □

4.2. **Isotonicity of the inverse demand correspondence.** 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 $\left(\delta \vee \delta'\right)_j = \max\left\{\delta_j, \delta'_j\right\}$. Theorem 2 expresses that $\tilde{\Sigma}^{-1}$ is isotone as a set-valued function in Veinott's (2005) strong set order.

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' in $\mathcal{D}_0$ such that $s \in \tilde{\Sigma}(\delta)$ 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 4.2. To 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. □

Before we examine the consequences of theorem 2, we explore what it 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), the conclusion of theorem 2 follows from Topkis' (1998) theorem. In this case, Chiong, Galichon, and Shum (2016) have shown that for* $s \in \mathcal{S}_0$

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

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*

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

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.* □

4.3. **The identified set is a lattice.** We now consider the consequences of theorem 2 for the structure of the identified set $\tilde{\Sigma}^{-1}(s)$. Specifically, we show that whenever it is nonempty, it is a connected lattice.[8] The literature on the estimation of discrete choice models has favored an approach based on imposing conditions guaranteeing invertibiliy of demand. In sharp contrast, our approach here imposes minimal assumptions, and provides analytical results and explicit algorithms for identification and estimation, *even when* the demand map (1.2) is not one-to-one and invertibility fails.

A key implication of Theorem 2 for identification is given in the next result:

Theorem 3. *Under assumption 1, if* $\tilde{\Sigma}^{-1}(s)$ *is nonempty, the following properties hold:*

(i) $\tilde{\Sigma}^{-1}(s)$ *is a lattice.*

(ii) $\tilde{\Sigma}^{-1}(s)$ *is a connected set.*

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

Part (i) of this result was originally proven in a discrete setting by Demange and Gale (1985) who show that the set of payoffs which ensures a stable allocation is a lattice with smallest and largest elements. The smallest element (or lower bound) corresponds to the unanimously most preferred stable allocation for the consumers ("consumer-optimal") and the unanimously least preferred stable allocation for the yogurts. Conversely, the largest element (or upper bound) corresponds to the unanimously least preferred stable allocation for the consumers and the unanimously most preferred stable allocation for the yogurts ("yogurt-optimal") [9]. Part (ii) of this result was proven in a discrete setting by David Gale, see Roth and Sotomayor (1992) and references therein.

The lattice structure of $\tilde{\Sigma}^{-1}(s)$ established in theorem 3 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. It is known since Gale and Shapley (1962) that the solution of a large class of two-sided matching problems have a lattice structure, whose extremal elements can be determined by two versions ("men propose first" or "women propose first") of the same algorithm. Similarly here, the matching algorithms that we shall describe below in section 6 will recover the extremal elements $\tilde{\delta}^{\min}(s)$ and $\tilde{\delta}^{\max}(s)$. Therefore, 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.

We summarize this discussion in the following result. In the sequel we shall introduce the domain of $\tilde{\Sigma}^{-1}$ as

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

and 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 3:

[9] In our setting it is the contrary, since $v_j = -\delta_j$

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 the precise connection is investigated in appendix A.1.

5. Existence and Uniqueness

The results in the previous section – especially those showing the lattice structure of the identified set – rely on the identified set being nonempty. In this section, we focus on this and several other related results. Readers interested primarily in the computational and empirical aspects of NARUM can proceed to Section 6.

First, we consider nonemptyness of the identified set: given s, when is the identified utility set $\tilde{\Sigma}^{-1}(s)$ nonempty? when is it restricted to a single point? When it is restricted to a single point, is it stable ? That is, given a small perturbation of P and s, are we led to a small perturbation in the solution utility vector δ?

In order to show that $\tilde{\Sigma}^{-1}(s)$ is nonempty, 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 1* (Stronger regularity of $\mathcal{U}$)**.** *Assume (Ω, P) is a Borel probability space for every $j \in \mathcal{J}_0$:*

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

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

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

Assumption 3 (Left and right behaviour)**.** *Assume that:*

(a) There is $a > 0$ such that $\mathcal{U}_{\epsilon j}(\delta)$ 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 $\Pr(\mathcal{U}_{\epsilon j}(\delta^*) > -a) < \eta$.

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

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 4. *Under assumptions 1*, and 3, $\tilde{\Sigma}^{-1}(s)$ is nonempty for all $s \in \mathcal{S}_0^{int}$.*

Note that assumption 2 is not needed for theorem 4, so that the theorem still applies in the case when the utility shocks are discrete, which is typically the case when simulation is used for computing the model (as in Section 6 below).

While in this paper we make use of theorem 4 to guarantee the existence of an identifying vector of demand shifters δ under very weak restrictions, this result is a contribution to matching theory per se. Indeed, it ensures the existence of a solution to the equilibrium transport problem, as introduced in Galichon (2015), definition 10.1.

5.1. **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 4. *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 $\mathbb{R}^J$.

The following examples should demonstrate that assumption 4 is actually quite natural.

Example 2.1 Continued. *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 continous, and* $Z_j = \delta_0 + \varepsilon_0 - \varepsilon_j$ *has a nonvanishing density over* $\mathbb{R}^J$ *as soon as* $(\varepsilon_0 - \varepsilon_j)$ *does.* □

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

5.2. **Consistency.** In practice, we will measure the vector of market share s up to sample uncertainty, and we shall 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 slighlty weaker than at the previous paragraph, we shall 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 converge to the unique δ in $\tilde{\Sigma}^{-1}(P, s)$.

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) 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 5 is stronger than assumption 4.

Theorem 6. *Under assumptions 1 and 5, assume that* P^n *and* s^n *converge weakly to* P *and* s*, respectively. By theorem 5,* $\tilde{\Sigma}^{-1}(P, s)$ *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.*

6. Computational implications

In the previous sections, the reformulation of discrete choice models (NARUM) as matching problem, has enabled us to establish results about identification. In this section, we show how it enables us to estimate these models.

The task of recovering the systematic utilities δ_j given observed choices $\{s_j\}_{j \in \mathcal{J}}$ can be done using the modified Gale-Shapley (deferred acceptance) algorithm described in Crawford and Knoer (1981). However, this algorithm is very slow and inefficient in our context.

Thus, we propose two new algorithms which have proved to be much faster. The first is inspired from Crawford and Knoer (1981) method and is based on the calculation of market shares for different vector of prices. The second is based on a reformulation of the stable matching problem as a Nash equilibrium in a supermodular game. We discuss the three methods in turn.

6.1. **Algorithm for identification 1: Deferred Acceptance.** In Crawford and Knoer's (1981) deferred acceptance algorithm, firms are proposing offers to workers, and the allocation which is found with this adjustment process yields lower bounds for salaries. Analogously, in our case, consumers are proposing offers to yogurts, and the allocation found is the upper bound for systematic utilities δ_j (or lower bound of the systematic utilities $-\delta_j$). Analogously, lower bounds for the utilities are found by reversing the roles: yogurts would propose to consumers.

For any given vector of market shares, the identified utilities can be recovered via discretization and simulation. We make the assumption that δ_j can only take integral values. In practice, this is not substantial as the grid of values for δ_j can be made arbitrarily fine, depending on the level of accuracy desired. We consider a matching scenario between N consumers and $N - q_0$ yogurts, with yogurt y having m_y copies. $n_0 = q_0$ consumers choose not to buy any yogurts (and get a reservation utility equal to $\mathcal{U}_{\varepsilon 0}(\delta_0)$ where δ_0 is fixed). A yogurt of brand j having $q_j = N \times s_j$ copies and n_0 consumers choose not to buy any yogurts. For the purpose of describing the algorithm, we need to distinguish between different *jars* of yogurt, which may however be of the same brand. This is because along the iterative path, prices for different jars of the same brand of yogurts may not have the same price. Therefore, let k index distinctive jars of yogurt. Also let j_k denote the brand of jar k. Hence, a jar k of brand j gives to a consumer the systematic utility δ_{j_k}.

R1. $\delta_{j_k} = \delta_{j_k}^{start}$. $\delta_{j_k}^{start}$ is chosen high enough so that no consumer prefers the outside option.

R2. Each consumer initially makes an offer to its favorite yogurt. Consumer ε, for example, makes an offer to the jar k, where k is a solution of the problem $\max_k \mathcal{U}_{\varepsilon j_k}(\delta_{j_k})$. Consumers may break ties at any time however they like.

R3. Each yogurt who receives one or more offers rejects all but one, which it tentatively accepts. Yogurts may break ties at any time however they like.

R4. Offers not rejected in previous periods remain in force. If yogurt k rejected an offer from consumer k in period $t-1$, $\delta_k^t = \delta_k^{t-1} - 1$; if not, $\delta_k^t = \delta_k^{t-1}$ Rejected consumers continue to make offers to their favorite jars.

R5. The process stops when no rejections are issued in some period. Jar then accept the offers that remain in force from the consumers they have not rejected.

The Crawford-Knoer algorithm presented here converges to the δ_j upper bound but is impractical computationally, as it is very slow. Thus, we propose two different algorithms in the following, which are new in the literature.

6.2. **Algorithm for identification 2: market shares adjustment with scaling.** The second algorithm is inspired by the former, but differs on two major points. First, it considers demand for brands and not for specific jars of yogurts, therefore, it decreases the systematic utilities δ_j of yogurt of the same brand, rather than adjusting each jar's utility separately. Second, it converges to the result by making several estimations which are first very loose and get more and more accurate.

As previously, we consider a matching scenario between N consumers and $N-n_0$ yogurts, with yogurt y having m_y copies. n_0 consumers choose not to buy any yogurts and get a reservation utility equal to $\mathcal{U}_{\varepsilon 0}(\delta_0)$ where δ_0 is fixed.

6.2.1. *Algorithm for the upper bound.* The algorithm is composed of two kind of rules. The first kind, rules R1 to R6, enables an approximation of the upper-bound. The second, R7, ensures convergence. Proofs of convergence can be found in the Appendix.

R1. At the beginning, all systematic utilities are set to a given value δ_j^0 higher than the upper bound. For $\left\{\delta_j^0\right\}_{j\in\mathcal{J}_0}$, the number of consumers who choose not to buy any yogurt is lower than at equilibrium m_0.

R2. At stage t, each consumer ε picks its preferred yogurt j based on maximizing $\max_j \mathcal{U}_{\varepsilon j}\left(\delta_j^t\right)$ breaking ties how she likes. If the number of consumers who choose the category j is strictly higher than the one at equilibrium m_j, the systematic utility of the yoghurt j is decreased by a given amount η_t: $\delta_j^t = \delta_j^{t-1} - \eta_t$.

R3. When the number of consumers who choose not to buy any yogurts is higher or equal to the one at equilibrium m_0, the algorithm stops. The larger η_t is, the faster this situation occurs.

R4. When the algorithm has stopped, at least some δ_j are lower or equal to the upper-bound, if not all. $\{\delta_j\}_{j\in\mathcal{J}}$ are increased by $\gamma \times \eta_t$ in order to have all δ_j higher than the upper-bound. Hence, γ should be big enough. Empirically, a good rule of thumb seems to be $\gamma = 2$.

R5. Once the $\{\delta_j\}_{j\in\mathcal{J}}$ have been increased, the algorithm starts again with a smaller value for η (for example, $\eta_{t+1} = 0.25 \times \eta_t$).

R6. The procedure (R2-R5) should be iterated until the necessary degree of accuracy is achieved (when η_t is small enough, let us call this value η_f).

Final step for convergence: At the end of the algorithm, there are two possibilities. (i) all δ_j are higher than the upper-bound. Applying R2-R3 again with a small enough η leads to the upper-bound (assuming a given degree of discretization). (ii) at least some δ_j are lower than the lower-bound. Applying R2 will either lead to a stable allocation δ_j and one systematic utility will not be decreased during the process or it will not converge to a stable allocation. The final rule for estimation R7 is based on these two remarks.

R7. Increase δ_j by η_f. Apply R2 and R3 with a chosen small value for η. If all systematic utilities have decreased and we have converged to a stable allocation, stop the algorithm. If not, start again with a higher initial value for δ_j (increase δ_j again by η_f for example).

6.2.2. *Algorithm for the lower bound.* In order to calculate the lower-bound, one could implement the same algorithm, but invert the roles of yogurts and consumers. However, this is not an efficient way to proceed since the problem is generally asymmetric: there are few brands of yogurts and a lot of different consumers. Therefore, the algorithm will 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.

In the following, a more efficient approach is developped. Instead of switching the roles of consumers and yogurts, we will instead directly adapt the upper-bound algorithm described above for the lower-bound. The rules are exactly the same as above, but with the following changes:

R1'. All systematic utilities are set to the lattice upper-bound.

R2'. At stage t, each consumer ε picks its preferred object j based on maximizing $\max_j \mathcal{U}_{\varepsilon j}\left(\delta_j^t\right)$ breaking ties however she likes. If the share of consumers who choose the category j is *higher than or equal to* the one at equilibrium m_j, the systematic utility of the yoghurt j is decreased by a given amount η_t: $\delta_j^t = \delta_j^{t-1} - \eta_t$.

R3'. δ_j stops decreasing when two conditions are met: (i) the number of consumers who choose not to buy any yogurts is strictly higher than at equilibrium, (ii) the number of consumers who choose j is strictly lower than at equilibrium. Note that the two conditions could be met at one point in the algorithm for a yogurt j, but if some $\delta_{k\neq j}$ decreases, then condition (ii) might not be met anymore for good j, δ_j should not be decreased in this case anyhow.

Final step for convergence: At the end of the algorithm, some if not all δ_j are below the lattice lower-bound. If it is the case, R7' will lead to the lower-bound. If some δ_j are above the lower-bound, then R7' will either lead to a stable allocation and at least one systematic utility will not have increased during the process or it will not converge to a stable allocation. Steps R7' and R8' are based on these principles.

R7'. Each consumer picks its preferred yogurt j based on maximizing $\max_j \mathcal{U}_{\varepsilon j}\left(\delta_j^t\right)$ breaking ties however she likes. If the share of consumers who choose the category j is *stricly lower than* the one at equilibrium m_j, the systematic utility of the yoghurt j is increased by a given amount η_f: $\delta_j^t = \delta_j^{t-1} + \eta_f$. The algorithm stops when the number of consumer who chooses the reference good is equal to the number of consumer at the equilibrium m_0.

R8'. If all systematic utilities have increased and we have converged to a stable allocation, stop the algoritm. If not, apply again R7' with lower value for δ_j.

Rules R1', R2', R3', R4, R5, R6, R7' and R8' make up the lower-bound algorithm.

6.3. **Numerical investigations.** In order to test our algorithms, we have made simulations with two different models. The first one is a model with linear transferable utility and the second one is a mixture vertical differentiation model. Both models are NARUM, and the second one is not point identified.

6.3.1. *Example 1: Linear transferable utility model.* In this model, both the intercept and slope in the utilities are choice-specific and random across consumers, hence, the systematic component δ_j do not enter linearly in the utility function, thus it is NARUM. The utility an individual i gets from the good j is defined as :

$$\mathcal{U}_{\varepsilon j}(\delta_j) = \alpha_{ij} + \delta_j \times \theta_{ij}$$

In our simulations, there are 100 consumers who choose between two different goods and an outside option. First, we have drawn α and θ from a standard uniform distribution. Then, we set different values for the δ_j ($\delta_{j_1} = 1$, $\delta_{j_2} = 1.5$). Since, we can calculate the utilities given by each good for each individual, we can recover the market shares of each good. Once the market shares have been computed, we use our algorithms to estimate the lattice bounds for the δ_j. The results are presented in Table 1.

For both models, the two algorithms converge quickly and accurately to the true values of the δ_j parameters we have set. The lower-bound and the upper-bound are very close which signals us that the theoretical model is probably point-identified (and it is).

6.3.2. *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.

	True	Nash		Market. Adjust.		Market
	δ_j	Lower b.	Upper b.	Lower b.	Upper b.	share
δ_{j_1}	1	0.99	1.00	0.99	1.00	48%
δ_{j_2}	1.5	1.49	1.53	1.49	1.53	19%
Ref. good	0	0	0	0	0	33%
Time	NA	0.16 sec	0.14 sec	0.44 sec	0.67 sec	NA

TABLE 1. Linear transferable utilities 100 consumers and 2 goods

There are three goods $y = 1, 2, 3$, and the unkown 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 $p_1^1 < p_2^1 < p_3^1$.
- 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:

$$s_y = 0.5(s_y^1 + s_y^2), \quad y = 1, 2, 3. \tag{6.1}$$

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[10]

$$\mathcal{U}_{\varepsilon_i,y}(\delta_y) \geq \mathcal{U}_{\varepsilon_i,y'}(\delta_{y'}) \Leftrightarrow \theta \geq \hat{\theta}_j(y, y') \equiv \min\left\{1, \max\left\{0, \frac{p_y^j - p_{y'}^j}{\delta_y - \delta_{y'}}\right\}\right\} \tag{6.2}$$

In this model, the δ's are partially identified. To see this, consider the two following cases:

[10]Note that, accoding to (6.2), 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.

(1) Good 3 has higher quality than 2, $\delta_1 < \delta_2 \leq \delta_3$. From (6.2), the market shares of store 1 are determined by $\hat{\theta}_1(2,1)$, $\hat{\theta}_1(3,1)$, and $\hat{\theta}_1(3,2)$ as follows:

$$\begin{aligned} s_1^1 &= \min\{\hat{\theta}_1(2,1), \hat{\theta}_1(3,1)\} \\ s_2^1 &= \max\{0, \hat{\theta}_1(3,2) - \hat{\theta}_1(2,1)\} \\ s_3^1 &= 1 - \max\{\hat{\theta}_1(3,2), \hat{\theta}_1(3,1)\} \end{aligned} \tag{6.3}$$

At store 2, no consumer buys good 2 because it is more expensive and has at most the quality of good 3. Hence, the market shares of store 2 are given by:

$$s_1^2 = \hat{\theta}_2(3,1); \quad s_2^2 = 0; \quad s_3^2 = 1 - \hat{\theta}_2(3,1) \tag{6.4}$$

Note that $s_1^1 = 0$ if $p_1^2 = p_3^2$ since good 3 has greater quality than good 1. The quality parameters $\{\delta_1, \delta_2, \delta_3\}$ that rationalize the observed market shares (s_1, s_2, s_3) solve (6.1) with the per-store market shares in (6.3) and (6.4).

(2) Good 2 has higher quality than 3, $\delta_1 < \delta_3 < \delta_2$. No consumer buys good 3 at store 1 since it is more expensive and has lower quality than good 2. Following the same reasoning as in case (1), we obtain the observed market shares as functions of the unkown parameters:

$$\begin{aligned} s_1 &= 0.5 * \hat{\theta}_1(2,1) + 0.5 * \min\{\hat{\theta}_2(2,1), \hat{\theta}_2(3,1)\} \\ s_2 &= 0.5 * [1 - \hat{\theta}_1(2,1)] + 0.5 * [1 - \max\{\hat{\theta}_2(2,1), \hat{\theta}_2(2,3)\}] \\ s_3 &= 0.5 * 0 + 0.5 * \max\{0, \hat{\theta}_2(2,3) - \hat{\theta}_2(3,1)\} \end{aligned} \tag{6.5}$$

Note that once again $p_1^2 = p_3^2 \Rightarrow \hat{\theta}_2(3,1) = 0 \Rightarrow s_1^2 = 0$. The quality parameters that rationalize the market shares (s_1, s_2, s_3) solve the system of equations (6.5).

Each case implies a different system of equations with possibly different solutions. This indeterminacy may imply multiple values of $\delta_1, \delta_2, \delta_3$ can rationalize the observed market shares s_1, s_2, s_3. In fact, we next provide a numerical example in which the set of quality parameters that rationalize the observed market shares is a continuum.

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

	True[11]	$N = 1000$		$N = 100$	
	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

Aggregate market shares: $(s_1, s_2, s_3) = (0.25, 0.25, 0.5)$

Prices: store 1 (1,2,3), store 2 (1,2,1)

Computed using market shares adjustment with scaling algorithm

TABLE 2. Mixture vertical differentiation: 3 goods

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

$$C = \{(\delta_2, \delta_3) : \delta_2 = 2, \delta_3 \in [1, 3]\} \tag{6.6}$$

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, are given in Table 2. The algorithm performs as expected, yielding the meet and the join.

7. Concluding remarks

In this paper we have explored the intimate connection between discrete-choice models and two-sided matching models, and used results from the literature on matching under imperfectly transferable utility to derive identification and estimation procedures for discrete-choice models based on the non-additive random utility specification.

Although it is commonplace to distinguish in the microeconomic literature between "one-sided" and "two-sided" demand problems, our results show that this distinction does not exist. Although yogurts do not derive utility from being consumed by one consumer or another one, the market adjusts in fashion which is observationally equivalent. 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[12] but is quite natural in the matching context, where "one-to-many" markets are commonplace – perhaps the most prominent and well-studied being the National Residents Matching Program for aspiring doctors in the United States (cf. Roth (1984)). We are exploring this connection in ongoing work.

References

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," .

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

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., 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," *The 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.

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.

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

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

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

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

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

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

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," *American Mathematical Monthly*, 69(1), 9–15.

Galichon, A. (2015): "Optimal Transport Methods in Economics," *Available at SSRN*.

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 Jr, A. S., and V. P. Crawford (1982): "Job matching, coalition formation, and gross substitutes," *Econometrica: Journal of the Econometric Society*, pp. 1483–1504.

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

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

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

Noldeke, G., and L. Samuelson (2015): "The implementation duality," .

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

Pakes, A., J. Porter, K. Ho, and J. Ishii (2011): "Moment Inequalities and their Applications," Discussion paper, working paper, University of Wisconsin.

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: Journal of the Econometric Society*, 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.

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

Topkis, D. (1998): "Supermodularity and Complementarity, 1998," .

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

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

Appendix A. Additional results

A.1. **Relation to Berry, Gandhi, and Haile's results.** The following result is a 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 $j \in \mathcal{J}$.

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, which implies that $\tilde{\delta}^{\min}(s) = \tilde{\delta}^{\max}(s)$ and that this function is inverse isotone. In contrast, in our setting, both $\tilde{\Sigma}$ and the injectivity of $\tilde{\Sigma}^{-1}$ 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).

Appendix B. Proofs

B.1. **Proof of theorem 1.**

Proof (a) From demand inversion to equilibrium matching: Consider $\delta \in \sigma^{-1}(s)$ 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 2.

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

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

In particular, for each open set B, one has

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

By Strassen's theorem (Strassen (1965), theorem 5, see also Galichon (2015) chap. 9.1), this implies that there is a probability distribution $\pi \in \mathcal{M}(P, s)$ such that $j \in J(\varepsilon)$ on the support of π. Hence π satisfy condition (i) in definition 2. But $j \in J(\varepsilon)$ implies $\mathcal{U}_{\varepsilon j}(\delta_j) = u(\varepsilon)$. Introducing $v(j) = -\delta_j$, $g_{\varepsilon j}(v(j)) = -\mathcal{U}_{\varepsilon_i j}(\delta_j)$, and $f_{\varepsilon j}(x) = x$, one has

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

for all (ε, j), with equality on the support of π. Hence, conditions (ii) and (iii) in definition 2 are met, and (π, u, v) is an equilibrium outcome.

<u>(b) From equilibrium matching to demand inversion</u>: Let (π, u, v) be an equilibrium matching in the sense of definition 2, where $f_{\varepsilon j}(x) = x$ and $g_{\varepsilon j}(y) = -\mathcal{U}_{\varepsilon_i j}(-y)$. Then letting $\delta = -v$, one has by condition (ii) that for any $\varepsilon \in \Omega$ and $j \in \mathcal{J}_0$,

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

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

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

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

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

Hence $\sigma(\delta) = s$, QED.

B.2. **Proof of proposition 1.**

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

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

Converse implication: Conversely, assume (B.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. ∎

B.3. **Proof of theorem 1'.** [TO BE COMPLETED NEED TO ADAPT A LITTLE THEOREM 1]

B.4. **Proof of theorem 2.**

Proof. Assume $s_j \leq s'_j$ for all $j \in \mathcal{J}$, and let δ and δ' in $\mathcal{D}_0$ such that $s \in \tilde{\Sigma}(\delta)$ and $s' \in \tilde{\Sigma}\left(\delta'\right)$. Let $u(\varepsilon) = \max_{j \in \mathcal{J}_0} U_{\varepsilon j}(\delta_j)$ and $u'(\varepsilon) = \max_{j \in \mathcal{J}_0} U_{\varepsilon j}\left(\delta'_j\right)$. 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 $s \in \tilde{\Sigma}\left(\delta \wedge \delta'\right)$: 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 all $A \subseteq \mathcal{J}_0$,

$$\sum_{j \in A} s_j \leq P\left\{\varepsilon \in \Omega : \exists j \in A,\ u(\varepsilon) = U_{\varepsilon j}(\delta_j)\right\}, \text{ and} \tag{B.2}$$

$$\sum_{j \in A} s'_j \leq P\left\{\varepsilon \in \Omega : \exists j \in A,\ u'(\varepsilon) = U_{\varepsilon j}\left(\delta'_j\right)\right\}. \tag{B.3}$$

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

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

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 (B.3) applied to $A = A^{>}$, one has

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

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

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

thus

$$\sum_{j \in A^{>}} s_j \leq P\left\{ \varepsilon \in \Omega^{>} : \exists j \in A,\ u^{\wedge}(\varepsilon) = U_{\varepsilon j}\left(\delta^{\wedge}_j\right) \right\}. \tag{B.5}$$

By (B.2) 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

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

By summation of (B.5) and (B.6), one obtains (B.4), 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 $B \subseteq \Omega$,

$$P(B) \leq \sum_{j \in \mathcal{J}_0} s_j 1\left\{\exists \varepsilon \in B : u(\varepsilon) = U_{\varepsilon j}(\delta_j)\right\}, \text{ and} \tag{B.7}$$

$$P(B) \leq \sum_{j \in \mathcal{J}_0} s'_j 1\left\{\exists \varepsilon \in B : u'(\varepsilon) = U_{\varepsilon j}\left(\delta'_j\right)\right\}. \tag{B.8}$$

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 $B \subseteq \Omega$,

$$P(B) \leq \sum_{j \in \mathcal{J}_0} s'_j 1\left\{\exists \varepsilon \in B : u^{\vee}(\varepsilon) = U_{\varepsilon j}\left(\delta_j^{\vee}\right)\right\}. \tag{B.9}$$

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

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

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 (B.7) applied to $B = B^{>}$, one has

$$P\left(B^{>}\right) \leq \sum_{j \in \mathcal{J}_0} s_j 1\left\{\exists \varepsilon \in B^{>} : u(\varepsilon) = U_{\varepsilon j}(\delta_j)\right\}; \tag{B.10}$$

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 (B.10) can be restricted to the elements of $\mathcal{J}_0^{>}$, which implies

$$\begin{aligned} P\left(B^{>}\right) &\leq \sum_{j \in \mathcal{J}_0^{>}} s_j 1\left\{\exists \varepsilon \in B : u(\varepsilon) = U_{\varepsilon j}(\delta_j)\right\} \\ &= \sum_{j \in \mathcal{J}_0^{>}} s_j 1\left\{\exists \varepsilon \in B : u^{\vee}(\varepsilon) = U_{\varepsilon j}\left(\delta_j^{\vee}\right)\right\}, \end{aligned} \tag{B.11}$$

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

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

Next, taking $B = B^{\leq}$ in (B.8) implies that

$$P\left(B^{\leq}\right) \leq \sum_{j \in \mathcal{J}_0} s'_j 1\left\{\exists \varepsilon \in B^{\leq} : u'(\varepsilon) = U_{\varepsilon j}\left(\delta'_j\right)\right\}; \tag{B.13}$$

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, (B.13) implies

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

By summation of (B.12) and (B.14), one obtains (B.9), and thus

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

∎

B.5. **Proof of theorem 3.**

Proof. Proof of (i): Let δ and δ' be two elements of $\tilde{\Sigma}^{-1}(s)$. Because $s \leq s$, theorem 2 implies $\delta \vee \delta' \in \tilde{\Sigma}^{-1}(s)$ and $\delta \wedge \delta' \in \tilde{\Sigma}^{-1}(s)$, QED.

Proof of (ii): Let $\bar{\delta}$ be the lattice upper bound of $\tilde{\Sigma}^{-1}(s)$. We need to show that for any δ^* in $\tilde{\Sigma}^{-1}(s)$, there is a continuous path connecting δ^* and $\bar{\delta}$ and remaining in $\tilde{\Sigma}^{-1}(s)$. To do this, introduce $L\left(\tilde{\delta}\right) = \tilde{\Sigma}^{-1}(s) \cap \left\{\delta \in \mathbb{R}^{\mathcal{J}} : \delta \leq \tilde{\delta}\right\}$, defined for $\tilde{\delta} \in \mathbb{R}^{\mathcal{J}}$, which is a lattice as soon as it is nonempty, and is a subset of $\tilde{\Sigma}^{-1}(s)$. Let

$$f\left(\tilde{\delta}\right) = \sup\left\{\delta \in \mathbb{R}^{\mathcal{J}} : \delta \in L\left(\tilde{\delta}\right)\right\}. \tag{B.15}$$

Because $L\left(\tilde{\delta}\right)$ is a lattice, one has $\left(f\left(\tilde{\delta}\right)\right)_i = \sup_{\delta \in L(\tilde{\delta})}\{\delta_i\}$. For the same reason $f\left(\tilde{\delta}\right) \in L\left(\tilde{\delta}\right)$ hence $f\left(\tilde{\delta}\right) \in \tilde{\Sigma}^{-1}(s)$ as soon as $L\left(\tilde{\delta}\right)$ is nonempty. Clearly, if $\delta \in \tilde{\Sigma}^{-1}(s)$, then $f(\delta) = \delta$. Now let $\delta_t = (1-t)\delta^* + t\bar{\delta}$, and $\hat{\delta}_t = f(\delta_t) \in \tilde{\Sigma}^{-1}(s)$. One has $\delta_0, \delta_1 \in$

$\tilde{\Sigma}^{-1}(s)$, so $\hat{\delta}_0 = \delta_0 = \delta^*$ and $\hat{\delta}_1 = \delta_1 = \bar{\delta}$. Further, by proposition 1, we have

$$L\left(\tilde{\delta}\right) \Leftrightarrow \begin{cases} \sum_{j\in B} s_j \leq P\left(\max_{j\in B} U_{\varepsilon j}(\delta_j) \geq \max_{j\in \mathcal{J}_0\setminus B} U_{\varepsilon j}(\delta_j)\right), \ \forall B \subseteq \mathcal{J}_0, \\ \delta_0 = 0 \\ \delta_j \leq \tilde{\delta}_j, \ \forall j \in \mathcal{J} \end{cases}$$

so f defined in (B.15) is continuous as a consequence of Berge's maximum theorem. Hence $\hat{\delta}_t = f(\delta_t)$ is a continuous path from δ^* to $\bar{\delta}$ in $L\left(\tilde{\delta}\right)$. ∎

B.6. **Proof of proposition 3.**

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. ∎

B.7. **Proof of theorem 4.** The proof is based on several lemmas.

Assumption 3 implies that for all $\eta > 0$ and $\nu > 0$ there is δ^* s.t. $\delta > \delta^*$ implies $\Pr(|X_\delta - X^*_\delta| > \nu) < \eta$.

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

$$\int \frac{\exp\left(\frac{\mathcal{U}_{\varepsilon j}(\underline{\delta}_j)}{T}\right)}{1+\exp\left(\frac{\mathcal{U}_{\varepsilon j}(\underline{\delta}_j)}{T}\right)} P(d\varepsilon) = s_j \tag{B.16}$$

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 , part (b) implies:

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

Next, by assumption 3, there exists $\underline{\delta}_j$ such that $\delta < \underline{\delta}_j$ implies $\Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta\right) > -a\right) \leq s_j/2$. Hence, for $\delta < \underline{\delta}_j$

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

and taking $T^* = a/\log\left(1/s_j - 1\right)$ if $\log\left(1/s_j - 1\right) > 0$, 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 $F_j^T\left(\delta\right) < s_j$.

Next, by assumption 3, there exists δ_j' such that $\delta > \delta_j'$ implies $\Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta\right) > 0\right) \geq 2s_j$. Then for $\delta > \delta_j'$,

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

As a result, we get:

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

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

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

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

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

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

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

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

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

We have

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

Next, by assumption 3, for all $b \in \mathbb{R}$, there exists δ_j^* such that $\delta > \delta_j^*$ implies $\Pr\left(\mathcal{U}_{\varepsilon j}(\delta) > b\right) \geq 1 - \eta/2$. Thus for $\delta_j > \delta_j^*$,

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

Choosing $b = -T \log(\eta/M)$ implies that the right hand-side is $\frac{1-\eta}{1+\eta} \geq (1-\eta)^2$. Because $\eta < 1 - \sqrt{s_j}$, $(1-\eta)^2 > s_j$, 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 $G_j^T(\delta)$ is decreasing with respect to $\delta_{j'}$ for any $j' \neq j$. Indeed,

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

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

$$G_j^T\left(\delta_j^{T,k}\right) \leq s_j, \tag{B.17}$$

From assumption 3, setting $\eta = s_0/4$ and $b = T^* \log\left(4/s_0 - 1\right)$, one has the existence of $\delta \in \mathbb{R}$ such that $\delta \geq \overline{\delta}_j$ implies $\Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta\right) < b\right) < \eta$.

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

$$\delta_j^k \leq \overline{\delta}_j \tag{B.18}$$

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

Proof. By summation of inequality (B.17) over $j \in \mathcal{J}$, one has

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

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

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

a contradiction. Thus inequality (B.18) holds. ∎

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

$$G_j^T\left(\delta_j^T; \delta_{-j}^T\right) = s_j. \tag{B.19}$$

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 (B.19). ∎

We can now deduce the proof of theorem 4.

Proof of theorem 4. 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

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

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

$$\left|\mathbb{E}_{\pi^{T_n}}\left[\mathcal{U}_{\varepsilon J}\left(\delta_J^{T_n}\right)-\mathcal{U}_{\varepsilon J}\left(\delta_J^0\right)\right]\right| \leq \nu \tag{B.20}$$

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

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

Combining (B.20) and (B.21), 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. ∎

B.8. **Proof of theorem 5.**

Proof. Let $\delta \in \tilde{\Sigma}^{-1}(s)$. 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. ∎

B.9. **Proof of theorem 6.** In order to prove this result, we shall need 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*

$$\bar{\delta}_j = \max_{(\delta_{-j}) \in \mathbb{R}^{\mathcal{J} \setminus \{j\}}} F(\delta_{-j}; P, s) \tag{B.22}$$

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*

$$F_{Bj}(\delta_j; P, \delta_{-j}) = P\left(\max_{k \in B} U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k) \geq \max_{k \in \mathcal{J}_0 \setminus (B \cup \{j\})} \left\{U_{\varepsilon j}^{-1} U_{\varepsilon k}(\delta_k), \delta_j\right\}\right). \tag{B.23}$$

Proof. By proposition 1, and theorem 3, if $\bar{\delta} = \sup \tilde{\Sigma}^{-1}(P; s)$, then

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

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

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

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

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

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

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

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

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

$$\left|F_{Bj}(\delta_j'; P, \delta_{-j}) - F_{Bj}(\delta_j; P, \delta_{-j})\right| \geq A\left|\delta_j' - \delta_j\right|. \tag{B.24}$$

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 5, $(X_B(\delta_{-j}), Y_B(\delta_{-j}))$ 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 (B.24) holds uniformly. ∎

We are now ready for the proof of the theorem.

Proof of theorem 6. Because assumption 5 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 (B.23), and E_{Bj} is defined by

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

hence

$$E_{Bj}\left(\delta_j^n;P,\delta_{-j}^n\right) - \epsilon_1^n - \epsilon_3^n \leq \sum_{k\in B} s_k \leq F_{Bj}\left(\delta_j^n;P,\delta_{-j}^n\right) + \epsilon_2^n + \epsilon_3^n$$

where

$$\begin{cases} \epsilon_1^n = \sup_{\|\delta\|\leq\rho}\left|E_{Bj}(\delta_j;P^n,\delta_{-j}) - E_{Bj}(\delta_j;P,\delta_{-j})\right| \\ \epsilon_2^n = \sup_{\|\delta\|\leq\rho}\left|F_{Bj}(\delta_j;P^n,\delta_{-j}) - F_{Bj}(\delta_j;P,\delta_{-j})\right| \\ \epsilon_3^n = \sum_{k\in B} s_k^n - \sum_{k\in B} s_k \end{cases},$$

hence $\eta_n := |\epsilon_3^n| + \max(|\epsilon_1^n|,|\epsilon_2^n|) \geq |F_B(\delta_i^n;\delta_{-i}) - F_B(\delta_i;\delta_{-i})| \geq A|\delta_i^n - \delta_i|$, and thus $|\delta_i^n - \delta_i| \leq \eta_n/A$, QED. ∎

Appendix C. Algorithms

C.1. **Algorithm 2 : market shares adjustment with scaling.**

C.1.1. *upper-bound.* The proof requires two lemmas.

Lemma 8. *If all δ_j are above or equal to the lattice upper-bound, then R2-R3 ensures convergence in finite time to the lower-bound.*

Proof. We will first show that systematic utilities will decrease as long as all systematic utilities are above or equal to the upper-bound and some are strictly above. Second, we will show that systematic utilities will stop increasing when they are equal to the upper-bound.

As long as all δ_j are above or equal to the upper-bound and some are strictly above, the number of people who consume some of these yogurts will be strictly higher than at equilibrium, and the systematic utilities of these yogurts will decrease according to R2. Indeed, if the number of consumers who consume each brand of yogurts is equal to its value at equilibrium, then the δ_j corresponds to a core allocation. This is absurd since we have supposed that some systematic utilities were above the lattice upper-bound. If the number of consumers who choose each brand of yogurt is lower or equal to its value at equilibrium, and strictly lower for some brands, then the number of consumers who choose the reference category is strictly higher than at equilibrium, which is not possible since demand is decreasing with the systematic utilities the other categories ($\sum_\varepsilon \mathbb{1}\left(\mathcal{U}_{\varepsilon 0}(\delta_0) \geq \mathcal{U}_{\varepsilon j'}(\delta'_j)_{j' \neq 0}\right)$ is decreasing in $\left\{\delta'_j\right\}_{j' \neq 0}$).

Second, if the discretization is fine enough, then δ_j will hit its upper bound. Then the number of consumers who choose this good $\hat{m}_y$ can not be strictly higher than at equilibrium as long as no $\left\{\delta'_j\right\}_{j' \neq j}$ is strictly higher than the lower bound since $\hat{m}_j = \sum_\varepsilon \mathbb{1}\left(\mathcal{U}_{\varepsilon j}(\delta_j) \geq \mathcal{U}_{\varepsilon j'}(\delta'_j)_{j' \neq j}\right)$ is increasing in $\left\{\delta'_j\right\}_{j' \neq j}$.

These arguments ensure that the systematic utilities of all yogurts will hit their upper-bound in a finite number of steps. ∎

However, this lemma is not sufficient to ensure the convergence of the algorithm since we can not guarantee that all systematic utilites are above the upper-bound in the first place. We need to have a rule to acknowledge this situation.

Lemma 9. *If at least one systematic utility δ_j is strictly lower than the lattice upper bound, then at least one systematic utility will not decrease during the process or the process will not converge to a core allocation.*

Proof. If at least one systematic utility is stricly lower than the lattice upper-bound, then two situations can occur. First, one systematic utility can be lower than the lower-bound.

In this case, the process will not converge to a core allocation. Second, if all systematic utilities are either higher than the lattice upper-bound or equal to a systematic utility in a core allocation, then the algorithm will converge to a core allocation, and one systematic utility will not increase. Indeed, if j' has the price δ'_j which corresponds to the stable allocation the least favorable for the consumers of all the initial value δ_j corresponding to a core allocation, then, following R2 all the other systematic utilities decrease until we reach the stable allocation corresponding to δ'_j (if several allocations corresponds to δ'_j, the algorithm converges to the most favorable one for the consumers). The same argument as for Lemma 10 can be used here. ∎

C.1.2. *lower-bound.* The proof requires two lemmas.

Lemma 10. *If all δ_j are below the lattice lower-bound, then R7' ensures convergence in finite time to the lower-bound.*

Proof. First systematic utilities will increase as long as all systematic utilities are lower or equal to the lower-bound and some are strictly lower. Second, systematic utilities will stop increasing when they are equal to the lower-bound. Proofs follows the same logic as for the upper-bound. If we assume some degree of discretization R7' will ensure convergence in a finite number of steps.

∎

However, this lemma is not sufficient to ensure the convergence of the algorithm since we can not guarantee that all systematic utilites are below the lower-bound in the first place. We need to have a rule to acknowledge this situation.

Lemma 11. *If at least one systematic utility δ_j is strictly higher than the lattice lower-bound, then at least one systematic utility will not increase during the process or the process will not converge to a core allocation.*

Proof. The same argument as for 9 can be used to make the proof. ∎