Empirical Welfare Economics[1]

Christopher P. Chambers
Department of Economics
Georgetown University
Christopher.Chambers@georgetown.edu

Federico Echenique
Department of Economics
UC Berkeley
fede@econ.berkeley.edu

September 27, 2022

[1]This paper is dedicated to the memory of Kim Border. We are grateful to audiences at the CUHK-HKU-HKUST Joint Theory Seminar, UC Berkeley, the University of Chicago, McGill University, the 2022 Social Choice and Welfare meetings, Stanford University, the Workshop on Applications of Revealed Preferences, and Roy Allen for detailed comments.

Abstract

Welfare economics relies on access to agents' utility functions: we revisit classical questions in welfare economics, assuming access to data on agents' past choices instead of their utilities. Our main result considers the existence of utilities that render a given allocation Pareto optimal. We show that a candidate allocation is efficient for some utilities consistent with the choice data if and only if it is efficient for an incomplete relation derived from the revealed preference relations and convexity. Similar ideas are used to make counterfactual choices for a single consumer, policy comparisons by the Kaldor criterion, and determining which allocations, and which prices, may be part of a Walrasian equilibrium.

1 Introduction

Consider a social planner facing a collection of agents in a classical resource allocation problem. Pareto optimality is characterized by the equality of agents' marginal rates of substitution, but to use this characterization our planner needs access to agents' utility functions. Suppose instead that the planner has access to a dataset consisting of a finite set of demand observations for each individual. The planner wants to know which allocations can be Pareto efficient for the collection of agents, given what she knows from the observed dataset. As a minimal discipline, she asks that there are monotone and convex preferences that are consistent with the data, and for which a given allocation is Pareto efficient.

Our main result provides a complete characterization of the allocations that can be Pareto efficient for the observed dataset. Our characterization parallels the definition of Pareto optimality, with an empirical domination relation standing in for unobservable utility comparisons. So the characterization says that there should be no dominating alternative allocation, where the notion of domination captures what can be inferred about agents' utilites from the dataset. In particular, the dataset defines a revealed preference relation. The revealed preference is, in general, incomplete; it does not compare all alternatives. Given revealed preference, we can speak of making further comparisons based on monotonicity, transitivity, and convexity. For example, if it is known that both x and y are revealed preferred to z, then $\frac{1}{2}(x + y)$ should also be at least as good as z. Further, imposing monotonicity allows for additional comparisons: if x is revealed preferred to z, and $w \geq x$, then w should also be preferred to z. Each such comparison can be further combined with transitivity in order to impose additional comparisons. All the inferences that we can make, using indirect revealed preference, convexity, and monotonicity, define what we call a domination relation for each individual agent. This domination relation is, in a sense, the "smallest" set of inferences we can make from the data by using rationality, convexity and monotonicity alone.

The domination relation is typically highly incomplete. Incompleteness results from the limitations in the information contained in the data, even when augmented by the consequences of assuming monotone and convex preferences. This is in contrast with the normative statements about incomplete preferences, as in the work of Ok (2002); Dubra et al. (2004); Eliaz and Ok (2006). Efficiency with respect to our relation is the same notion as is used in the matching literature, where the incomplete relation is typically the stochastic dominance relation on a set of lotteries induced by a linear order on the set of degenerate outcomes. See e.g. Bogomolnaia and Moulin (2001); McLennan (2002); Abdulkadiroğlu and Sönmez (2003); Manea (2008); Carroll (2010); Bogomolnaia and Heo (2012); Hashimoto et al. (2014); Aziz et al. (2015); Doğan and Yıldız (2016).

Our main result says, moreover, than the concavity of utility has no empirical content above the convexity of the underlying preferences, at least for the question of deciding whether an allocation could be Pareto optimal. Our main result says that an allocation is Pareto optimal for some monotone and (explicitly) quasiconcave utilities if and only if it is Pareto optimal for some monotone and concave utilities that are consistent with the data.

The paper actually uses the domination relation, and related concepts, to address a host of related questions in welfare economics. We start from individual welfare comparisons, and ask for counterfactual comparisons that may be inferred from individual-level consumption data. In particular, given data from one consumer, and two new bundles x and y, we ask when one can infer that the utility of x is greater than that of y, for all rationalizing concave utilities. The exercise follows Varian (1982), and is related to the literature on demand bounds; see e.g. Blundell et al. (2007, 2008, 2015); Allen and Rehbeck (2020b,a). Our answer depends on a notion of empirical domination that is closely related to the notion behind our result on Pareto domination. There is, again, an empirically defined partial order among consumption bundles that captures all the comparisons that may be inferred from the dataset and the hypotheses of monotonicty, transitivity and concavity (note: not simply convexity as in the efficiency question we described aboive).

Next, we turn to collective welfare comparisons. Aside from our main result on Pareto optimal allocations, which we have already described, we consider the Kaldor criterion: whether an economic policy decision can be defended on the grounds that those who benefit from the policy could compensate those who lose (Kaldor, 1939; Hicks, 1939). Again the idea of domination gives us an answer, and serves to rule out whether demand data validates a policy decision.

Our methods can be used to discuss the testable implications of Walrasian equilibrium, in the spirit of Brown and Matzkin (1996). Given demand data, we characterize the prices that could be Walrasian equilibrium prices. In the General Equilibrium literature, the famous Sonnenschein-Mantel-Debreu theorem (Shafer and Sonnenschein, 1982; Chambers and Echenique, 2016) can be read as saying that the are no restrictions on the sets of prices that may be equilibrium prices. Brown and Matzkin show that data on prices and endowments (observations "on the equilibrium manifold") may be used to refute the theory, but they do not characterize the prices that are consistent with the theory. Our result provides such a characterization, when the data assumed are individual-level consumption data.

Finally, we turn our attention to the existence of a representative consumer. There are well-known impossibility results that rule out a representative consumer, unless the income distribution is severely restricted. Our result shows that if agents' preferences may be inferred from the data, and the distribution allowed to be chosen as part of the rationalization exercise, then representative consumers may be obtained very generally. We think of this result as a caveat on the idea of endogenizing the income distribution to enable a representative consumer.

Related Literature.

The theory of efficiency in classical economic environments without completeness is studied in many works; a few of these include Shafer and Sonnenschein (1975); Gale and Mas-Colell (1975, 1977); Fon and Otani (1979); Weymark (1985); Rigotti and Shannon (2005) Bewley (2002), and Bewley et al. (1987).

Also related are concepts of testing whether certain allocations can be equilibria of a given economy. Brown and Matzkin (1996) is a canonical reference. In that paper, the authors check whether a collection of candidate objects could be equilibria of a given economy. Results in this literature usually focus on establishing a list of polynomial inequalities that must be satisfied in order for the data to be rationalizable—these inequalities are analogous to the "Afriat inequalities" of rational consumer behavior. In showing that a particular rationalization problem reduces to one of verifying whether a solution exists to a list of polynomial inequalities establishes that these problems are decidable, in an algorithmic sense. See also Brown and Shannon (2000); Bossert and Sprumont (2002); Kubler (2003); Carvajal et al. (2004); Carvajal (2004); Bachmann (2004, 2006b,a); Brown and Calsamiglia (2007); Brown and Kubler (2008); Carvajal (2010); Cherchye et al. (2011); Carvajal and Song (2018) for testable implications of related environments. Some of these investigate efficiency directly: Bossert and Sprumont (2002) discuss how the core correspondence varies (for fixed preferences) as endowments vary. Bachmann (2006b) considers an environment in which collections of endowments and consumption bundles (but not prices) are observed. His Proposition 5 establishes that Pareto efficiency has essentially no testable content in this environment, even if all preferences are represented by strictly concave and continuously differentiable utilities.[1]

As mentioned, when it comes to welfare comparisons, what these papers primarily do is provide an analogue of the result of Afriat (1967), whereby rationalizability is equivalent to the satisfaction of a set of inequalities. In contrast, our work differs in two respects: first, we provide an economic characterization of whether a given bundle could possibly be efficient—our characterization is more analogous to the characterization of rationality via absence of cycles (also discussed by Afriat (1967), and termed "Generalized Axiom of Revealed Preference" by Varian (1982)). We take as the starting point of our proof a collection of "Afriat inequalities" that must be satisfied, and use these

[1]The idea is that a common linear preference renders every allocation efficient. Then perturb each agent's utility a bit to ensure strict concavity and smoothness.

to uncover a dual system of linear inequalities that we can interpret — they have concrete economic meaning — and deliver a condition in terms of the domination relation.

Second, we focus on a single, candidate allocation. In so doing, we are able to come up with a formulation of the problem in which the equations we must solve are *linear*. This formulation is what allows us to leverage well-known duality techniques. Were we to ask the same question for multiple candidate allocations, the problem would be polynomial. Importantly, there may be two candidate allocations, each of which are possibly efficient, but which cannot possibly both be efficient at the same time.

We are not the first to study representative consumers in a revealed preference framework. Cherchye et al. (2009) consider household preference aggregation in a model with a collective public good, and Cherchye et al. (2016) establish an empirical counterpart to the Gorman aggregation result. Their focus is on empirically understanding two sources of aggregation: household bargaining and linear Engel curves. Our result focuses instead on endogenous income distribution, as in Samuelson (1956), but not necessarily with the presence of a social welfare function.[2] So, the income distribution is allowed to depend on the aggregate budget but not necessarily with a goal toward optimizing some type of social welfare. Our result establishes the inherent weakness of not restricting the income distribution.

2 The model

Basic definitions and notational conventions.

We use the following notational conventions: For vectors $x, y \in \mathbf{R}^n$, $x \leq y$ means that $x_i \leq y_i$ for all $i = 1, \ldots, n$; $x < y$ means that $x \leq y$ and $x \neq y$; and $x \ll y$ means that $x_i < y_i$ for all $i = 1, \ldots, n$. The set of non-negative vectors in $\mathbf{R}^n$ is denoted $\mathbf{R}^n_+$, and the set of vectors that are strictly positive

[2] In general an endogenous income distribution which ensures rational aggregate behavior need not arise from maximization of a social welfare function. See Dow and da Costa Werlang (1988).

in all components is $\mathbf{R}^n_{++}$. A function $f : A \subseteq \mathbf{R}^n \to \mathbf{R}$ is *weakly monotone increasing*, or *non-decreasing*, if $f(x) \leq f(y)$ when $x \leq y$; and *monotone increasing*, if it is weakly monotone increasing and $f(x) < f(y)$ when $x \ll y$. We often just write "increasing."

A function $u : \mathbf{R}^n_+ \to \mathbf{R}$ is *concave* if, for all $x, y \in \mathbf{R}^n_+$ and $\lambda \in (0, 1)$,

$$u(\lambda x + (1 - \lambda)y) \geq \lambda u(x) + (1 - \lambda)u(y);$$

and *quasiconcave* if, for all $x, y \in \mathbf{R}^n_+$ and $\lambda \in (0, 1)$,

$$u(\lambda x + (1 - \lambda)y) \geq \min\{u(x), u(y)\}.$$

The function u is *explicitly quasiconcave* if it is quasiconcave and, for all $x, y \in \mathbf{R}^n_+$ and $\lambda \in (0, 1)$, $u(x) \neq u(y)$ implies that

$$u(\lambda x + (1 - \lambda)y) > \min\{u(x), u(y)\}.$$

Observe that explicit quasiconcavity is a behavioral property, meaning a property of the preference relation represented by u, and that it is weaker than concavity. Indeed, explicit quasiconcavity is only a minor strengthening of quasiconcavity; it is weaker than strict quasiconcavity ($u(\lambda x + (1 - \lambda)y) > \min\{u(x), u(y)\}$ for all $\lambda \in (0, 1)$), which corresponds to strict convexity of preferences. Strict quasiconcavity rules out that indifference curves contain any flat regions (i.e contain any line segments), but flat regions are consistent with explicit quasiconcavity, outside of some rather pathological examples. Perhaps explicit quasiconcavity is best known because it ensures that local maxima are global maxima, for which quasiconcavity alone does not suffice (see Theorem 192 in Border (2015)).

Definitions from welfare economics.

An agent is defined through a preference relation on $\mathbf{R}^m_+$, which we represent throughout by a utility function $u : \mathbf{R}^m_+ \to \mathbf{R}$.[3] The elements of $\mathbf{R}^m_+$ are called *consumption bundles*. Given a finite set of agents N, an *allocation* is a vector $\bar{x} = (\bar{x}_i)_{i\in N} \in \mathbf{R}^{mN}_+$.[4] If each agent is endowed with a utility function u_i, an allocation $\bar{y}$ *Pareto dominates* the allocation $\bar{x}$ if $u_i(\bar{y}_i) \geq u_i(\bar{x}_i)$ for all i, with a strict inequality for at least one agent. An allocation $\bar{x}$ is *Pareto optimal* if there is no allocation satisfying

$$\sum_{i\in N} \bar{y}_i = \sum_{i\in N} \bar{x}_i$$

that Pareto dominates it.

Next we turn to a criterion for comparing allocations based on the principle that winners may compensate the losers. The idea is that those who gain in moving from one allocation to the other may compensate those who lose with the move in allocations. Let $\bar{x}$ and $\bar{y}$ be two allocations. Say that $\bar{x}$ *weakly Kaldor dominates* $\bar{y}$ if there is no allocation $\bar{z}$ with $\sum_i \bar{z}_i \leq \sum_i \bar{y}_i$ that Pareto dominates $\bar{x}$. The idea is that if $\bar{x}$ does not weakly dominate $\bar{y}$, then there is a way of re-assigning (whence losers are compensated by winners) the aggregate bundle $\sum_i \bar{y}_i$ in a way that Pareto dominates $\bar{x}$ (see Chapter 5 in Graaff (1967) for a discussion of the Kaldor criterion).

An *exchange economy* is a tuple $E = (u_i, \omega_i)_{i\in N}$, where $i \in N$ is the set of agents in the economy, and each agent is endowed with a utility function u_i and an *endowment vector* $\omega_i \in \mathbf{R}^m_+$. A *Walrasian equilibrium* in E is a pair $((x_i)_{i\in N}, p)$ for which 1) $\sum_i x_i = \sum_i \omega_i$ (markets clear); and 2) for all $i \in N$, $p \cdot x_i = p \cdot \omega_i$ and $u_i(x_i') > u_i(x_i)$ implies that $p \cdot y > p \cdot \omega_i$.

Given endowment vectors $\omega_i \in \mathbf{R}^m_+$ for a set of agents N, we say that $\bar{x} = (\bar{x}_i)_{i\in N} \in \mathbf{R}^{mN}_+$ is an *allocation of* $(\omega_i)_{i\in N}$ when $\sum_i \bar{x} = \sum_i \omega_i$.

[3]So we restrict attention to continuous preference relations, but given that preferences are only constrained to rationalize a finite dataset, continuity is without loss of generality.

[4]One should think of an allocation $\bar{x}$ as "allocating" the aggregate bundle $\sum_{i\in N} \bar{x}_i$ among the agents in N.

Data and rationalizability.

A pair $(p, x) \in \mathbf{R}_+^{m+m}$ is an *observation*, and should be interpreted as the datum that the consumption bundle $x \in \mathbf{R}_+^m$ was chosen from the budget set $\{y \in \mathbf{R}_+^m : p \cdot y \leq I\}$ in which the income, or budget, is $I = p \cdot x$. A (possibly empty) finite list of observations $\{(p^k, x^k)\}_{k=1}^K$ is termed an *individual dataset.* N is a finite set of individuals. A *group dataset* is a collection of individual datasets, one for each $i \in N$. So, $D_i = \{(p_i^k, x_i^k)\}_{k=1}^{K_i}$ denotes an individual dataset for individual i, and $\{D_i : i \in N\}$ is a group data set.

An individual dataset is *rationalizable* if there is an increasing utility function $u_i : \mathbf{R}_+^m \to \mathbf{R}$ for which for all k, $u_i(x) > u_i(x_i^k)$ implies $p_i^k \cdot x > p_i^k \cdot x_i^k$. In this case, we say that u_i *rationalizes* the individual dataset (or that it is a *rationalizing* utility, when the dataset is implied). Similarly, we say that a group dataset is *rationalizable* if each individual dataset is rationalizable.

In our paper, we insist that rationalizing utilities be monotone increasing. Clearly, some structure must be assumed on utilities, or any data becomes rationalizable by a constant utility. The most common approach is to impose local non-satiation, and then resort to Afriat's theorem which says that one may without loss of generality assume a rationalizing utility that is both increasing and concave. Thus monotonicity, but more importantly concavity, comes for free in the case of an individual agent's observed consumption behavior.

Revealed preference theory requires the use of two binary relations. The *direct revealed preference* of agent i is denoted by $\succeq_i^R$, and defined by $x \succeq_i^R y$ if $x \geq x_i^k$ for some k, and $p_i^k \cdot x_i^k \geq p_i^k \cdot y$, or if $x = y$. The *direct strict revealed preference* of agent i is denoted by $\succ_i^R$, and defined by $x \succ_i^R y$ if

$$x \gg x' \succeq_i^R y, \text{ or } x \geq x_i^k \text{ and } p_i^k \cdot x_i^k > p_i^k \cdot y,$$

for some x' or observation k. These definitions of revealed preferences are slightly unusual, in that they already incorporate the expectation of a monotone preference, and symmetry is built-in.[5] Observe that $\succ_i^R \subseteq \succeq_i^R$.

[5]See Chambers and Echenique (2009) and Nishimura et al. (2017) for such "composi-

The *indirect revealed preference* $\succeq_i^I$ is defined as the transitive closure of $\succeq_i^R$. The *indirect revealed strict preference* $x \succ_i^I y$ obtains when there is a finite chain $x = z_1 \succeq_i^R \ldots \succeq_i^R z_L = y$, where at least one instance of $\succeq_i^R$ is $\succ_i^R$.

A dataset satisfies the *Generalized Axiom of Revealed Preference (GARP)* if there is no $x, y \in \mathbf{R}_+^m$ such that $x \succeq_i^I y$ while $y \succ_i^I x$.

3 Results

We consider counterfactual welfare comparisons. Given data on individual consumption, we seek to characterize which counterfactual (i.e. unobserved) welfare conclusions may be drawn on the basis of what can be inferred about agents' preferences from the data. For individual agents, we want to evaluate unobserved bundles. For a group of agents, the welfare comparisons are about the possible Pareto optimality of some allocation, or consistency with the Kaldor criterion. The same ideas allow us to understand the possible (again counterfactual) Walrasian equilibrium prices, and when a representative agent is possible.

All proofs are relegated to Section 5.

3.1 Individual welfare

We begin by discussing individual welfare conclusions that may be drawn from a single agent's consumption dataset. Aside from the intrinsic merit of these results, they serve to introduce some of the ideas we use later in our (main) results on collective welfare.

Our first result asks when we can say that one bundle is unambiguously better than another, given what the data tell us about the agent. Specifically, given an individual dataset $\{(x^k, p^k) : 1 \leq k \leq K\}$, and two bundles $\bar{x}$ and $\bar{y}$, when is $\bar{x}$ ranked above $\bar{y}$ for all utility functions compatible with the data?

The answer turns out to depend on the revealed preference relation inferred

tions" of the revealed preference relation with the partial order on consumption bundles. It is easy to see that Afriat's theorem remains true under our definition of revealed preference.

from the consumer's choices. Say that $\bar{x}$ *bests* $\bar{y}$ if $\bar{x}$ can be written as a convex combination of bundles z^l, where for each l $z^l \succeq^I \bar{x}$, or $z^l \succeq^I \bar{y}$, with at least one occurrence of the latter. Say that $\bar{x}$ *strictly bests* $\bar{y}$ if it weakly bests it, and one of the revealed-preference comparisons is strict ($\succ^I$ for $\succeq^I$).[6]

It is easy to see that if $\bar{x}$ strictly bests $\bar{y}$, then it is ranked above $\bar{y}$ by any rationalizing concave and monotone increasing utility function. Indeed, if $\bar{x} = \sum_l \lambda_l z^l$ is as above, then for any concave, increasing, rationalizing utility:

$$\begin{aligned} u(\bar{x}) &\geq \sum_l \lambda_l u(z^l) \\ &\geq \alpha u(\bar{x}) + (1-\alpha) u(\bar{y}) \end{aligned}$$

with $\alpha < 1$ because at least one of the z^l corresponds to a comparison with $\bar{y}$. Given that at least one inequality is strict we conclude that $u(\bar{x}) > u(\bar{y})$.

Our first result says that the condition is not only sufficient for the conclusion, but also necessary.

Theorem 1. *Let $\{(x^k, p^k) : 1 \leq k \leq K\}$ be an individual dataset and $\bar{x}, \bar{y} \in \mathbf{R}^m_+$ be two bundles. Then $u(\bar{x}) > u(\bar{y})$ for all concave and monotone rationalizing u if and only if $\bar{x}$ strictly bests $\bar{y}$.*

Besting is useful to compare two counterfactual bundles, but we shall need a somewhat different concept for our results on collective choices. Our next result is a warm-up for the analysis of collective welfare because it will involve the same variation of "besting," which we term "domination." The question is not about ranking two consumption bundles, but instead we are given an unobserved bundle $\bar{x}$, and wish to know when there exists a rationalizing utility for which this new bundle is at least as good as anything that was observed in the data.

[6] A bundle $\bar{x}$ strictly bests itself when it is incompatible as a choice with the existing dataset. This means that there is no price $\bar{p}$ at which $\bar{x}$ could be demanded, and for which the resulting dataset (obtained by adding $(\bar{x}, \bar{p})$ to the dataset) is rationalizable. If the dataset is rationalizable, however, we may choose $\bar{p}$ that supports the upper contour set of a (without loss, concave) rationalizing utility at $\bar{x}$. Adding the resulting observation to the dataset preserves its rationalizability.

Say that a bundle y *weakly dominates* $\bar{x}$ if it is a convex combination of some collection z^l of bundles, $1 \leq l \leq L$, such that, for each l, $z^l \succeq^I \bar{x}$.

A bundle y *strictly dominates* $\bar{x}$ for agent i if it weakly dominates it and, moreover, if in the defining convex combination there is l with $z^l \succ^I \bar{x}$.

Theorem 2. *Let $\{(x^k, p^k) : 1 \leq k \leq K\}$ be an individual dataset and $\bar{x} \in \mathbf{R}^m_+$ an arbitrary bundle. There exists an increasing and explicitly quasiconcave rationalizing utility for which $u(\bar{x}) \geq \max\{u(x^k) : 1 \leq k \leq K\}$ if and only if, once we add $\bar{x} \succeq^R x^k$ for all k to the revealed preference relation, as well as as well as $x^k \succeq^R \bar{x}$ when $p^k \cdot (\bar{x} - x^k) \leq 0$ and $x^k \succ^R \bar{x}$ when $p^k \cdot (\bar{x} - x^k) < 0$, we have*

1. *GARP is satisfied.*
2. *There is no bundle $y \leq \bar{x}$ that strictly dominates $\bar{x}$.*

In contrast with Theorem 1, which wanted something to be true of every (concave, increasing) utility, Theorem 2 asks about the existence of a rationalizing utility with a certain property. The latter sort of result is, of course, most conclusive when the condition fails, and thus certifies that the property is incompatible with any rationalizing utility. Finally, observe that Theorem 2 only asks utilities to be explicitly quasiconcave. The same will be true of our main result below.

3.2 Collective welfare

Our main result characterizes the allocations that are efficient for some utility functions (with the requisite properties) that are consistent with a group dataset.

An allocation $\bar{y}$ *empirically dominates* the allocation $\bar{x}$ if $\sum_i \bar{y}_i \leq \sum_i \bar{x}_i$ while $\bar{y}_i$ weakly dominates $\bar{x}_i$ for all i and strictly dominates it for at least one i. Observe the parallelism with the notion of Pareto domination: Given increasing utility functions $(u_i)_{i \in N}$ we may say that an allocation $\bar{y}$ Pareto dominates $\bar{x}$ if $\sum_i \bar{y}_i \leq \sum_i \bar{x}_i$, while $u_i(\bar{y}_i) \geq u_i(\bar{x}_i)$ for all i, and $u_i(\bar{y}_i) > u_i(\bar{x}_i)$

for at least one i. Theorem 3 implies that empirical domination really is the empirical counterpart to Pareto domination.

Theorem 3. *Let* $\{(x_i^k, p_i^k) : 1 \leq k \leq K_i\}$, *for* $i \in N$, *be a rationalizable group dataset, and* $\bar{x}$ *an allocation. The following statements are equivalent:*

1. *There are increasing, and explicitly quasiconcave, rationalizing utilities for which* $\bar{x}$ *is Pareto efficient.*

2. *There are increasing, and concave, rationalizing utilities for which* $\bar{x}$ *is Pareto efficient.*

3. *The allocation* $\bar{x}$ *is not empirically dominated by any other allocation.*

The theorem provides a characterization of the allocation that could be efficient, for some monotone and convex preferences of the agents (with the minor strengthening of convexity implied by explicit quasiconcavity). The role of the unobserved utility functions in the definition of Pareto domination is taken by the observable empirical domination relations.

An important message in Theorem 3 is that concavity comes for free. In consumer theory, explicit quasiconcavity is a behavioral property: a property of the consumer's preference relation. Concavity of utility is a cardinal property of the consumer's utility function. Similar to Afriat's theorem, Theorem 3 essentially says that in a world of agents with convex preferences, concavity has no testable implications when it comes to detecting efficient allocations.

Empirical domination ensures the existence of a common supporting price at the allocation $\bar{x}$, essentially the equality of marginal rates of substitution for a collection of rationalizing utilities. If we additionally require that this price supports the *Scitovsky contour* at $\bar{x}$, then the ideas behind Theorem 3 can be used to provide an empirical basis for the Kaldor criterion:[7]

Corollary 4. *Let* $\{(x_i^k, p_i^k) : 1 \leq k \leq K_i\}$, *for* $i \in N$, *be a rationalizable group dataset. Let* $\bar{x}$ *and* $\bar{y}$ *be allocations. There are increasing, concave,*

[7] Given utilities (u_i), the *Scitovsky contour* at $\bar{x}$ is the set $S(\bar{x}) = \{\sum_i z_i : u_i(z_i) \geq u_i(\bar{x}_i) \text{ for all } i \in N\}$. If a price q supports all individual upper contour sets at $\bar{x}$ and $q \cdot \sum_i \bar{y}_i < q \cdot \sum_i \bar{x}$, then $\sum_i \bar{y}_i \notin S(\bar{x})$.

rationalizing utilities for which $\bar{x}$ weakly Kaldor dominates $\bar{y}$ if there is no allocation $(\bar{z}_i)$ that weakly dominates $\bar{x}_i$ for all i, and strictly dominates it for at least one i, and a scalar $\kappa \geq 0$, for which

$$\sum_i \bar{z} \leq \sum_i \bar{x}_i + \kappa(\sum_i \bar{y}_i - \sum_i \bar{x}_i)$$

Observe that Corollary 4 only offers a sufficient condition for Kaldor domination. When the condition holds, then we may say that there are rationalizing utilities for which a switch from $\bar{x}$ to $\bar{y}$ could not be defended on the basis of the Kaldor criterion.

Given Theorem 3, one may use the Second Welfare Theorem to decentralize a possibly efficient allocation $\bar{x}$ by means of taxes and subsidies. But one may also want to know when $\bar{x}$ is a potential Walrasian allocation without any transfers. Suppose then that we have access to individual endowments (ω_i), for which $\sum_i \omega_i = \sum_i \bar{x}_i$, and we want to know if there are prices q for which $(\bar{x}, q)$ constitutes a Walrasian equilibrium of the exchange economy defined by the endowments and some rationalizing utilities.

Say that a bundle $\bar{y}_i$ *ω_i-dominates* $\bar{x}_i$ if $\bar{y}_i$ is the convex combination of bundles z_i^l where, for each l, either $z_i^l = \omega_i$ or $z_i^l \succeq_i^I \bar{x}_i$. Say that a bundle $\bar{y}_i$ *strictly ω_i-dominates* $\bar{x}_i$ if $\bar{y}_i$ ω_i-dominates $\bar{x}_i$ and one of the inequalities in the convex combination is strict: so there is l with $z_i^l \succ_i^I \bar{x}_i$.

Theorem 5. *Let $\{(x_i^k, p_i^k) : 1 \leq k \leq K_i\}$, for $i \in N$, be a rationalizable group dataset. Suppose given a collection $(\omega_i)_{i\in N}$ of endowments, and an allocation $(\bar{x}_i)_{i\in N}$ of $(\omega_i)_{i\in N}$. There exists a price vector q, and increasing, concave, rationalizing utilities $(u_i)_{i\in N}$ so that $(q, (\bar{x}_i))$ is a Walrasian equilibrium of $(u_i, \omega_i)_{i\in N}$ if and only if there is no allocation $(\bar{y}_i)_{i\in N}$ of the endowments so that 1) $\bar{y}_i$ ω_i-dominates $\bar{x}_i$ for all i, and 2) strictly ω_i-dominates it for some i.*

3.3 Walrasian equilibrium

Motivated by the Sonnenschein-Mantel-Debreu theorem, which implies that nothing can be said about the sets of prices that can be Walrasian equilibrium prices, Brown and Matzkin (1996) famously argued that general equilibrium theory *has testable implications* for data on prices and individual-level incomes. Brown and Matzkin's result relies on the decidability of certain systems of polynomial equations, but they do not provide a characterization of the data that are consistent with Walrasian equilibrium.[8] Here we shall provide such a characterization, but under somewhat different assumptions. We take as given a group dataset, a collection of individual endowments, and a price vector that is a candidate for equilibrium price. Our result provides a condition that describes when the price can be a Walrasian equilibrium price.

Formally, we have access to a group data set, and we are given 1) agents' endowments $(\omega_i)_{i\in N}$, and 2) a proposed Walrasian equilibrium price $\bar{p}$.[9] We want to know if there is an allocation $(\bar{x}_i)$ such that $((\bar{x}_i), \bar{p})$ constitutes an Walrasian equilibrium in the exchange economy $(u_i, \omega_i)_{i\in N}$, for some collection of rationalizing utilities $(u_i)_{i\in N}$.

Note that for any given price $\bar{p}$ we can say whether an observed bundle x_i^k would be affordable at the budget defined by $\bar{p}$ and endowments ω_i: this will happen when $\bar{p}\cdot x_i^k \le \bar{p}\cdot\omega_i$. So we can think of $\bar{p}$ as a "partial" observation, to be added to the data of each individual agent, which describes a new price and budget, but not a chosen consumption bundle. We may say that, whatever a consumer chooses to buy at this budget, it would be revealed preferred to x_i^k if $\bar{p}\cdot x_i^k \le \bar{p}\cdot\omega_i$, and strictly revealed preferred to x_i^k if $\bar{p}\cdot x_i^k < \bar{p}\cdot\omega_i$.

Now, with $\bar{p}$ in hand, such revealed preference comparisons should be added to those already defined from the existing data. Then we may take the transitive closure of the revealed preference relations thus augmented by $\bar{p}$, and say that a bundle x is *empirically worse* than consumption at prices $\bar{p}$ if, whatever would be consumed at $\bar{p}$ would be indirectly revealed preferred to x. Similarly we may say that a bundle x is *strictly empirically worse* than consumption at prices $\bar{p}$ if the revealed preference relation is strict. Let L_i be the set of observations for which the consumption bundles are empirically worse than $\bar{p}$.

[8] They do provide such a characterization, in terms of what they call the Weak Axiom of Revealed Equilibrium, for the special case of $N = 2$ and $K_i = 2$.

[9] A similar result is possible if we assume given individual incomes instead of endowments. The same is true of Brown and Matzkin (1996).

We adopt the following notation: $I_i = \bar{p} \cdot \omega_i$ is i's income when prices are $\bar{p}$ and her endowment ω_i; $I_i^k = p_i^k \cdot x_i^k$ is agent i's implied income in observation k, and $\bar{\omega} = \sum_i \omega_i$ is the economy's aggregate endowment. We say that $\bar{p}$ is *consistent* with the group dataset if there is a choice for individual consumption at prices $\bar{p}$ that does not violate GARP. It is possible to provide a characterization of consistent prices, essentially along the lines of our results in Section 3.1. In the statement of the theorem, a and b are the first two letters of the alphabet; they are disjoint from $\bigcup_i L_i$.

Theorem 6. *Consider a rationalizable group dataset, a consistent price $\bar{p}$, and endowments $(\omega_i)_{i\in N}$. There are increasing, concave, rationalizing utilities (u_i), and consumption bundles $\bar{x}_i$, for $i \in N$, so that $((\bar{x}_i), \bar{p})$ constitutes a Walrasian equilibrium of the exchange economy $(u_i, \omega_i)_{i\in N}$ if and only if there is no price $q^* \in \mathbf{R}_+^m$ and probability μ_i on $L_i \cup \{a, b\}$ such that*

1. $\mathbf{E}_{\mu_i}\tilde{p}_i \leq q^*$ *for all* i,
2. *and* $\sum_i \mathbf{E}_{\mu_i}\tilde{I}_i > q^* \cdot \bar{\omega}$,

where $\tilde{p}_i$ and $\tilde{I}$ are random price and incomes that equal, respectively, p_i^k and I_i^k on $k \in L_i$, $\bar{p}$ and I_i on a, and 0 on b.

The condition in the theorem means that there is a "social," or common, price q^* that all agents agree is undesirable, but makes total income cheaper: meaning that q^* is bad because it makes goods more expensive than at an average of either $\bar{p}$ or at prices that are already revealed to be worse than $\bar{p}$, and at the same time makes aggregate endowment (= total income) cheaper than the average observed or proposed income. More specifically, suppose that v_i is agent i's indirect utility function. Then $\mathbf{E}_{\mu_i}\tilde{p}_i \leq q^*$ for all i implies that $\mathbf{E}_{\mu_i} v(\tilde{p}_i) \geq v(\mathbf{E}_{\mu_i}\tilde{p}_i) \geq v_i(q^*)$, as v_i is convex and nonincreasing. The condition in Theorem 6 says that, to rule out that $\bar{p}$ is an equilibrium price, the unfavorable price q^* would still price aggregate endowment below the agents' aggregate expected income.

3.4 Representative consumer

We now turn to the existence of a representative consumer. It is well-known that a representative consumer is impossible under other than very stringent assumptions: Antonelli's Theorem (Antonelli, 1886) and Gorman's Theorem (Gorman, 1953) deliver clear impossibility results when one insists on the representative consumer being valid for all price vectors and individual budgets (see for example Shafer and Sonnenschein (1982)). The literature has therefore turned to situations where the income distribution is endogenously determined by some efficient allocation rule. Our next result looks at this question when all we know about consumers comes from data on their consumption choices.

For convenience we assume that all observed prices are the same. The more important substantive assumption is the existence of a "small" agent, who always consumes less than the aggregate bundle in every observation. Our result says that endogenizing an income distribution in this setting enables the existence of a representative consumer quite generally.

Theorem 7. *Let $D_i = \{(x_i^k, p_i^k) : 1 \leq k \leq K_i\}$, for $i \in N$, be a group dataset with the property that $K = K_i$ and $p_i^k = p^k$ for all i, and that, for some agent i^*, $x_{i^*}^l < \sum_i x_i^k$ for all k, l. Let $D_a = \{(\sum_i x_i^k, p^k) : 1 \leq k \leq K\}$ be the associated aggregate dataset. Then the datasets D_a and D_i, for all $i \in N$, are rationalizable if and only if there are increasing, concave, rationalizing utilities u_i for each agent $i \in N$, and v for the aggregate dataset D_a, so that for any price vector $p \in \mathbf{R}_+^m$ and income $I > 0$ there are $(x_i) \in \mathbf{R}_+^{mN}$ such that*

1. $\sum_i x_i \in \text{argmax}\{v(z) : z \in \mathbf{R}_+^m \text{ and } p \cdot z \leq I\}$
2. $x_i \in \text{argmax}\{u_i(z) : z \in \mathbf{R}_+^m \text{ and } p \cdot z \leq p \cdot x_i\}$

In Theorem 7, $p \cdot x_i$ should be read as agent i's endogenous income. So the property that $x_i \in \text{argmax}\{u_i(z) : z \in \mathbf{R}_+^m \text{ and } p \cdot z \leq p \cdot x_i\}$ means that i is optimizing by choosing x_i at prices p and income set to $I_i = p \cdot x_i$.

One interpretation of Theorem 7 comes from the property of rationalizability. If we are interested in aggregation, it is natural to consider a situation where a group data set *and* the resulting aggregate dataset D_a are rationalizable. Theorem 7 describes what may be inferred theoretically from such a situation.

4 Remarks

They key to our results is an observation based on Afriat's theorem, which says that an individual dataset $\{(p_i^k, x_i^k) : 1 \leq i \leq K_i\}$ is rationalizable if and only if there is a solution $U_i^k, \lambda_i^k > 0$ to the following system of linear "Afriat inequalities:"[10]

$$U_i^l \leq U_i^k + \lambda_i^k p_i^k \cdot (x_i^l - x_i^k).$$

The observation is that we may normalize such a solution so that $\lambda_i^{k^*} = 1$ for some specific observation k^*. As a result we obtain that system that remains linear, even if the prices $p_i^{k^*}$ at this particular observation were unknown.

With this observation in hand, we can now approach a problem like that in Theorem 3. For the allocation $\bar{x}$ to be Pareto optimal, agents' utilities would need to have a common supporting price q at $\bar{x}_i$. The existence of such a price q may be added to the above system of inequalities as if it were a new observation. Assuming that the corresponding value of λ has been normalized to 1, the system is still linear. See Bachmann (2004) or Bachmann (2006b) for related constructions. Now the work in proving the theorem amounts to interpreting the dual linear system.

The results obtained in Section 3 exemplify the power of our approach, but there are also clear limits. Given a dataset, one may ask a related question for a *collection* of allocations: whether there exists a single economy capable of generating all such allocations as Pareto efficient ones. It is natural to conjecture that there is such an economy if and only if each of the allocations is undominated. This conjecture turns out to be false, as shown by the following example:

[10] See Chambers and Echenique (2016) for a discussion of Afriat's theorem and this system of linear inequalities.

Example 1. *Let* $N = \{1, 2\}$*, and suppose there are two commodities, so that* $m = 2$*. Individual* 1 *has an empty individual dataset. Individual* 2 *has four observations:* $(p_2^1, x_2^1) = ((2, 1), (1, 2))$*,* $(p_2^2, x_2^2) = ((2, 1), (0, 4))$*,* $(p_2^3, x_2^3) = ((1, 2), (2, 1))$*, and* $(p_2^4, x_2^4) = ((1, 2), (4, 0))$*.*

Now, suppose we want to consider the allocations $\bar{x}_1^1 = (1, 0)$*,* $\bar{x}_2^1 = (0, 4)$*, and* $\bar{x}_1^2 = (0, 1)$*,* $\bar{x}_2^2 = (4, 0)$*. Observe that because individual* 1 *has an empty individual dataset, each of these allocations are possibly efficient by Theorem 3. On the other hand, they cannot both be efficient for the same economy. To understand why, observe that if* q^1 *supports* x_2^1*, then* $q^1 \cdot (0, 4) \leq q^1 \cdot (1, 2)$*, as the individual data set for individual* 2 *is rational. If* $q^1(2) = 0$ *(the second coordinate of* q^1*), then this inequality is obviously strict as* $q^1 \geq 0$*.*

So, if $q^1(2) = 0$*, we conclude that* $q^1 \cdot (1, 2) - q^1 \cdot (0, 4) > 0$*, so that* $q^1 \cdot (1, -2) > 0$*, from which we conclude* $q^1 \cdot (1, -1) > 0$*, or* $q^1 \cdot x_1^1 > q^1 \cdot x_1^2$*. Similarly, if* $q^1(2) > 0$*, then we know* $q^1 \cdot (1, -2) \geq 0$*, so that (as* $q^1(2) > 0$*),* $q^1 \cdot x_1^1 > q^1 \cdot x_1^2$*.*

So, $q^1 \cdot x_1^1 > q^1 \cdot x_1^2$*; symmetrically,* $q^2 \cdot x_1^2 > q^2 \cdot x_1^1$*. These inequalities obviously cannot simultaneously hold for a rational decision maker.*

In our discussion, we reduced the problem of testing whether an allocation $\bar{x}$ could be efficient to the question of the existence of a supporting price q. Were we to ask that multiple allocations be efficient, we would need a different supporting prices for each such allocation, but more to the point, the scale factors could differ across individuals, thus rendering the system nonlinear. In other words, we would need different λ for the different allocations, and the normalization would no longer help us.

So there are obvious limits to our approach, but there are also additional applications that we have not exhausted. One of these is envy-freeness. Suppose given a group dataset, and consider the existence of rationalizing utilities that render some proposed allocation $\bar{x}$ envy-free: meaning rationalizing utilities (u_i) with the property that $u_i(\bar{x}_i) \geq u_i(\bar{x}_j)$ for all $i, j \in N$. Our methods, based on working through the dual of *augmented* system of Afriat inequalities, provide an answer to this question.

A sketch of the solution follows: the trick is to add supporting prices for each agent at the proposed consumption of other agents in the allocation $\bar{x}$. The normalization idea keeps the system linear, and we just need to include utility values $u_{i,j}$ for i's utility at the bundle intended for j:

1. For all $i \in N$ and all $k, l \in \{1, \dots, K_i\}$ for which $p_i^l \cdot (x_i^k - x_i^l) \leq 0$, we have $u_i^k \leq u_i^l + \lambda_i^l p_i^l \cdot (x_i^k - x_i^l)$.
2. For all $i, j \in N$ and all $k \in \{1, \dots, K_i\}$ for which $p_i^k \cdot (\overline{x}_j - x_i^k) \leq 0$, we have $u_{i,j} \leq u_i^k + \lambda_i^k p_i^k \cdot (\overline{x}_j - x_k^i)$.
3. For all $i, j \in N$ and all $k \in \{1, \dots, K_i\}$, $u_i^k \leq u_{i,j} + p_{i,j} \cdot (x_k^i - \overline{x}_j)$.
4. For all $i, j, h \in N$, $u_{i,j} \leq u_{i,h} + p_{i,h} \cdot (\overline{x}_j - \overline{x}_h)$.
5. For all $i, j \in N$, $u_{i,i} \geq u_{i,j}$.

We omit the details, but hope that it is clear how to proceed on the basis of this system.

5 Proofs

5.1 Proof of Theorem 3

We begin with the following lemma, which is stated in Chambers and Echenique (2016), Remark 3.6.

Lemma 8. *Let $i \in N$. Suppose that for all $k \in \{1, \dots, K_i\}$, there are $u_i^k \in \mathbf{R}$ and $\lambda_i^k > 0$ for which for all $k, l \in \{1, \dots, K_i\}$ satisfying $p_i^k \cdot (x_i^l - x_i^k) \leq 0$, we have*

$$u_i^l \leq u_i^k + \lambda_i^k p_i^k \cdot (x_i^l - x_i^k).$$

Then the individual dataset $\{(p_i^k, x_i^k)\}_{k=1}^{K_i}$ is rationalizable.

Proof. Suppose that the condition in the statement of the Lemma is satisfied. Define the pair of binary relations $x_i^k \succeq_i^R x_i^l$ if $p_i^k \cdot (x_i^l - x_i^k) \leq 0$ and $x_i^k \succ_i^R x_i^l$ if $p_i^k \cdot (x_i^l - x_i^k) < 0$.

A *cycle* is a finite list $x_i^{l_1} \succeq_i^R x_i^{l_2} \succeq_i^R \dots x_i^{l_a} \succ_i^R x_i^{l_1}$. We claim that there can be no cycle. For, if there were, then we would have:

$$u_i^{l_{j+1}} - u_i^{l_j} \le \lambda_i^{l_j} p_i^{l_j} \cdot (x_i^{l_{j+1}} - x_i^{l_j}),$$

for all $j = 1, \dots, a-1$ and

$$u_i^{l_1} - u_i^{l_a} \le \lambda_i^{l_a} p_i^{l_a} \cdot (x_i^{l_1} - x_i^{l_a}).$$

Reading addition of indices as modulo a, observe that

$$0 = \sum_{j=1}^{a} (u_i^{l_{j+1}} - u_i^{l_j}) \le \sum_{j=1}^{a} \lambda_i^{l_j} p_i^{l_j} \cdot (x_i^{l_{j+1}} - x_i^{l_j}) < 0.$$

The first equality is by telescoping, the weak inequality by summing the original inequalities, and the strict inequality because of the right hand sides of the original inequalities are nonpositive (and at least one strictly negative). So, we arrive at a contradiction and there can be no cycle. Conclude by Afriat's Theorem (Afriat, 1967; Chambers and Echenique, 2016) that the individual dataset is rationalizable. □

Now we proceed with the proof of the theorem.

First, that (1) implies (3) follows because if u_i are rationalizing monotone and explictly quasiconcave utilities, then $z_i \succeq_i^I \bar{x}_i$ implies $u_i(z_i) \ge u_i(\bar{x}_i)$, and $z_i \succ_i^I \bar{x}_i$ implies $u_i(z_i) > u_i(\bar{x}_i)$. So when y_i is a convex combination of bundles $z_i^l \succeq_i^I \bar{x}_i$ we must have that $u_i(y_i) \ge u_i(\bar{x}_i)$ by quasiconcavity of utility. Moreover, if $z_i^l \succ_i^I \bar{x}_i$ for some l then we obtain $u_i(y_i) > u_i(\bar{x}_i)$ by explicit quasiconcavity. In all, then, when y_i dominates $\bar{x}_i$ for all agents, and strictly dominates for at least one agent, we have that $\bar{x}$ is Pareto dominated for the rationalizing utilities.

Second, it is obvious that (2) implies (1). So we focus our attention on showing that (3) implies (2). (Indeed our argument shows that (2) and (3) are equivalent.) Suppose then that (3) is satisfied. We will demonstrate that there exists some $q \in \mathbf{R}^m_{++}$ so that, for all $i \in N$, the individual dataset given by $\{(p_i^k, x_i^k)\}_{k=1}^{K_i} \cup \{(\overline{x}_i, q)\}$ is rationalizable. This then implies (by Afriat's Theorem) the existence of a concave, increasing utility function for which for all $y \in \mathbf{R}_+^m$ satisfying $q \cdot y \leq q \cdot \overline{x}_i$, we have $u_i(y) \leq u_i(\overline{x}_i)$, and consequently that $u_i(y) > u_i(\overline{x}_i)$ implies $q \cdot y > q \cdot \overline{x}_i$. Consequently, it also follows that $u_i(y) \geq u_i(\overline{x}_i)$ implies $q \cdot y \geq q \cdot \overline{x}_i$, by continuity and monotonicity of u_i. It then follows that $\overline{x}$ is efficient for these utility indices.[11]

The proof relies on a homogeneous Theorem of the Alternative: see Border (2020).

The content of Afriat's Theorem is that for each $i \in N$ and $k \in \{1, \ldots, K_i\}$, there is u_i^k and $\lambda_i^k > 0$ for which for all $k, l \in \{1, \ldots, K_i\}$,

$$u_i^k \leq u_i^l + \lambda_i^l p_i^l \cdot (x_i^k - x_i^l).$$

What we would now like to find are additional unknown parameters. Namely, for each $i \in N$, a scalar $\overline{u}_i \in \mathbf{R}$ and $q \in \mathbf{R}^m$. The vector q is required to be common to all individuals and will reflect the common prices supporting the hypothesized efficient allocation $\overline{x}$.

Our task is then to find $q \in \mathbf{R}^m$, and for each $i \in N$, a real number $\overline{u}_i \in \mathbf{R}$, and for each $i \in N$ and $k \in \{1, \ldots, K_i\}$, $u_i^k \in \mathbf{R}$ and $\lambda_i^k \in \mathbf{R}$ for which the following linear inequalities are satisfied:

1. For all $i \in N$ and all $k, l \in \{1, \ldots, K_i\}$ for which $p_i^k \cdot (x_i^l - x_i^k) \leq 0$, we have $u_i^l \leq u_i^k + \lambda_i^k p_i^k \cdot (x_i^l - x_i^k)$.
2. For all $i \in N$ and all $k \in \{1, \ldots, K_i\}$, $u_i^k \leq \overline{u}_i + q \cdot (x_i^k - \overline{x}_i)$.
3. For all $i \in N$ and all $k \in \{1, \ldots, K_i\}$, for which $p_i^k \cdot (\overline{x}_i - x_i^k) \leq 0$, we have $\overline{u}_i \leq u_i^k + \lambda_i^k p_i^k \cdot (\overline{x}_i - x_i^k)$.
4. For all $i \in N$ and all $k \in \{1, \ldots, K_i\}$, $\lambda_i^k > 0$.
5. $q \geq 0$ and $q \neq 0$.

[11] If not, then there is $\overline{y}$ for which $\sum_i \overline{y}_i = \sum_i \overline{x}_i$ and for all $i \in N$, we have $u_i(\overline{y}_i) \geq u_i(\overline{x}_i)$, with inequality strict for some $j \in N$, implying $\sum_i q \cdot \overline{y}_i > \sum_i q \cdot \overline{x}_i$, a contradiction.

The inequalities can be represented in matrix notation. We display part of the matrix below, as the matrix itself is quite large. The matrix below displays four horizontal blocks. The first two correspond to vectors corresponding to weak inequalities, the latter two to strict. This matrix has, for each agent i, $2(K_i + 1)$ columns, and an additional m columns; in total the number of columns is $m + \sum_i (2K_i + 1)$. Observe that, in the matrix written below, the column labelled by q actually represents m columns; for example, $\mathbf{1}_{m'}$ is an indicator function of the dimension $m' \in \{1, \ldots, m\}$.

As to rows, the matrix has, for each agent i, one row for each ordered pair (l, k) where $l, k \in \{1, \ldots, K_i\}$, $k \neq l$, and $p_i^k \cdot (x_i^l - x_i^k) \leq 0$. When agent i is understood, the row is labeled (l, k), as in the displayed matrix below. Continuing with the rows for agent i, there are also three rows for each k: one labeled by $(k, *)$, one by $(*, k)$ and one by k. The row labeled (k, l) for agent i is meant to capture inequality (1): there is a 1 in the column k for agent i, a -1 in column l, and $p_i^k \cdot (x_i^l - x_i^k)$ in the column for k among the second set of K_i columns. The rest of the entries in that row are zero. In a similar vein, the rows labeled by $(k, *)$ and $(*, k)$ are there to encode the inequalities in (2) and in (3). The row labeled k is meant to capture the basic positivity constraint (4), and has a one in column k, among the second collection of K_i columns.

Finally, the matrix has a collection of rows $m + 1$ that are not specific to any agent and seek to capture (5). There is then one column for each $m' \in \{1, \ldots, m\}$ (labelled $(*, m)$), expressing the nonnegativity of q, and a row asserting that $\sum_{m'=1}^{m} q(m') > 0$; the row labelled M.

Because this matrix is large, we only show certain portions of it. The rows listed in the matrix have zeroes everywhere for every remaining column.

$$
\begin{array}{c|ccccccccc|ccccc|c}
 & 1 & \cdots & k & \cdots & l & \cdots & K_i & \cdots & * & 1' & \cdots & k' & \cdots & K_i' & q \\
(l,k) & 0 & \cdots & 1 & \cdots & -1 & \cdots & 0 & \cdots & 0 & 0 & \cdots & p_i^k \cdot (x_i^l - x_i^k) & \cdots & 0 & 0 \\
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & 0 \\
(*,k) & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & \cdots & -1 & 0 & \cdots & p_i^k \cdot (\overline{x}_i - x_i^k) & \cdots & 0 & 0 \\
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & 0 \\
(k,*) & 0 & \cdots & -1 & \cdots & 0 & \cdots & 0 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 & x_i^k - \overline{x}_i \\
\hline
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & 0 \\
(*,m') & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \mathbf{1}_{m'} \\
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & 0 \\
\hline
M & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \mathbf{1}_{\{1,\ldots,m\}} \\
\hline
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & 0 \\
k & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \vdots & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & 0
\end{array}
$$

We are searching for a vector in $m + \sum_i (2K_i + 1)$ dimensional real space which, when multiplied with this matrix to yield a linear combination of its columns, results in a vector whose coordinates in the first two horizontal blocks are nonnegative, and in the last two are strictly positive. Such a vector would represent a solution to the system of inequalities (1)-(5). This is the system to which we will apply a duality result.

By Motzkin's transposition theorem (a version of the theorem of the alternative, see Theorem 47 in Border (2020)) there is no solution to the set of inequalities (and consequently to the enumerated list of inequalities above) if and only if there is, for each row of the matrix, a nonnegative weight, where for some row corresponding to a strict inequality (either in the third or fourth horizontal block), one of the weights is strict, for which the weighted sum of rows is the zero vector.

So, let us suppose by means of contradiction that there is no solution to the linear system. Therefore, there exists a solution to the dual system. Interpret the solution as a collection of weights on the rows of the matrix. For the rows corresponding to agent $i \in N$ (any row except the one labelled M), we let $\xi_i^A \geq 0$ denote the weight for the row labelled by A. For example, in the row of the above matrix labelled (l, k), $\xi_i^{(l,k)}$ is the associated weight. We let $\xi^M \geq 0$ be the weight associated with row M (which is common to all $i \in N$), and we let $\xi^{(*,m')} \geq 0$ be the weight associated with row $(*, m')$.

The matrix has a special structure. Observe that, restricted to the first $\sum_i (K_i + 1)$ block of columns on the left, and the rows labeled (k, l), $(k, *)$, or $(k, *)$ for some agent (and some k, l), the matrix becomes the incidence matrix of a graph with vertexes that can be identified with these $\sum_i (K_i + 1)$ columns. So each vertex is identified with a pair (i, k), of an agent and an observation $k \in \{1, \ldots, K_i\}$, or with a pair $(i, *)$ for the hypothesized efficient bundle. An edge goes from a node (i, k) to (i, l) when $p_i^k \cdot (x_i^l - x_i^k) \leq 0$. An edge goes from $(i, *)$ to (i, k) when $p_i^k \cdot (\overline{x}_i - x_i^k) \leq 0$. An edge always goes from (i, k) to $(i, *)$.

Now, the solution to the dual, when restricted to the incidence submatrix, provides a non-negative linear combination of rows that equals the null vector. The Poincaré-Veblen-Alexander theorem (Berge, 2001) claims that for any non-negative weighted sum of incidence vectors of a directed graph which is zero, there is a collection of positively oriented cycles in the graph, each cycle being associated with a weight, and the total weight ascribed to an incidence vector is the sum of all weights associated to cycles in which the incidence vector appears. Here, a cycle includes no repetitions of nodes.

Because the individual dataset $\{(p_i^k, x_i^k)\}_{k=1}^{K_i}$ is rationalizable, we may assume without loss of generality that every such cycle involves an edge of the type connecting (i, k) to $(i, *)$. This is because otherwise, along all elements of the cycle, rationalizability implies that $p_i^{k_j} \cdot (x_i^{k_{j+1}} - x_i^{k_j}) = 0$, and thus the weighted sum of vectors across that cycle is zero. Removing them does not affect the total weighted sum of rows.

Let us now represent the cycles associated with agent $i \in N$ by $\mathcal{C}_i$, as described, each of them comes with a weight $\mu(c) \geq 0$. What we just claimed is that for each $c \in \mathcal{C}_i$, there is some $k \in \{1, \ldots, K_i\}$ and an edge connecting (i, k) to $(i, *)$. This implies, in particular, that $x_i^k \succeq_i^I \bar{x}_i$. To see why, let the cycle be written via a sequence of nodes: $(i,*), (i,k_1), \dots, (i,k_l = k), (i,*)$. Because $(i,*)$ is connected to (i,k_1) by an edge, it means that $p_i^{k_1} \cdot (\bar{x}_i - x_i^{k_1}) \leq 0$, so that $x_i^{k_1} \succeq_i^R \bar{x}_i$; similarly, $x_i^{k_{j+1}} \succeq_i^R x_i^{k_j}$ for all $j = 1, \dots, l-1$. Consequently, by definition, $x_i^k \succeq_i^* \bar{x}_i$.

What we have just claimed is that if $\xi_i^{(k,*)} > 0$, it must be that $x_i^k \succeq_i^I \bar{x}_i$.

Now, again by Motzkin's transposition theorem, one of the following must be true: either $\xi^M > 0$, or there is $i \in N$ and $k \in \{1, \dots, K_i\}$ for which $\xi_i^k > 0$.

Let us consider each of the two cases in turn.

Case 1: There is a dual solution with $\xi^M > 0$.

The only columns for which row M are nonzero are the last m columns. Rows of type $(*, m')$ add (potentially) non-negative terms to these last m columns. Since the weighted sum of rows equals zero, it follows that

$$\sum_i \sum_{k=1}^{K_i} \xi_i^{(*,k)} (x_i^k - \overline{x}_i) = -\sum_{m'=1}^{m} \xi^{*,m'} \mathbf{1}_{m'} - \xi^M \mathbf{1}_{1 \dots, m} \ll 0. \tag{1}$$

In other words, for each $i \in N$ and each $k \in \{1, \dots, K_i\}$, there is a number $\theta_i^k \geq 0$ for which

$$\sum_i \sum_{k=1}^{K_i} \theta_i^k (x_i^k - \overline{x}_i) \ll 0,$$

where by the preceding discussion, $\theta_i^k > 0$ implies $x_i^k \succeq_i^I \bar{x}_i$. Furthermore, there is $i \in N$ and $k \in \{1, \dots, K_i\}$ for which $\theta_i^k > 0$, since equation (1) is strictly negative in every coordinate.

Without loss of generality (since the system is homogeneous), we may assume that $\sup_{i \in N} \sum_{k=1}^{K_i} \theta_i^k = 1$.

For each $i \in N$, let $\theta_i^0 = 1 - \sum_{k=1}^{K_i} \theta_i^k$. Then

$$\sum_i (\theta_i^0 \bar{x}_i + \sum_k \theta_i^k x_i^k) = \sum_i (\bar{x}_i + \sum_k \theta_i^k (x_i^k - \bar{x}_i)) \ll \sum_i \bar{x}_i.$$

So we can define

$$\bar{y}_i = \theta_i^0 \bar{x}_i + \sum_{k=1}^{K_i} \theta_i^k x_i^k.$$

for all $i \neq 1$. Observe that $\bar{y}_i$ is a convex combination of $\bar{x}_i \succeq_i^I \bar{x}_i$ (by definition), and $x_i^k \succeq_i^I \bar{x}_i$. If $\theta_1^0 > 0$, choose $y_1' \gg \bar{x}_1$ so that $\bar{y}_1 = \theta_1^0 y_1' + \sum_{k=1}^{K_1} \theta_1^k x_1^k$ and $y_1' \succ_1^I \bar{x}_1$; otherwise choose $y_1^{k^*} \gg x_1^{k^*}$ so that $\bar{y}_1 = \theta_1^0 \bar{x}_1 + \sum_{k=1}^{K_1} \theta_1^k x_1^k + \theta_1^{k^*}(y_1^{k^*} - x_1^{k^*})$ and $y_1^{k^*} \succ_1^I x_1^{k^*}$. Either way the allocation $\bar{y}_i$ weakly dominates $\bar{x}_i$ all agents, and strictly dominates it for agent 1.

Case 2: There is a dual solution with $\xi_i^k > 0$.

This means that there is $i \in N$ and $k \in \{1, \ldots, K_i\}$ for which $\xi_i^k > 0$. Fix such an $i^* \in N$ and a $k^* \in \{1, \ldots, K_i\}$. Because $\xi_M = 0$ is possible, we may only conclude in this case that $\sum_i \sum_{k=1}^{K_i} \xi_i^{(*,k)}(x_i^k - \overline{x}_i) \leq 0$.

On the other hand, we may conclude, since $\xi_{i^*}^{k^*} > 0$, that there is also $l \in \{1, \ldots, K_{i^*}\}$ with $\xi_{i^*}^{(l,k^*)} > 0$ and $p_{i^*}^{k^*} \cdot (x_{i^*}^l - x_{i^*}^{k^*}) < 0$; or in other words, $x_{i^*}^{k^*} \succ_i^R x_{i^*}^l$. In particular, the edge (i^*, k^*) to (i^*, l) belongs to some $c \in \mathcal{C}_i$, which has a corresponding $\xi_{i^*}^{(*,k)} > 0$; we may conclude then that $x_{i^*}^k \succ_{i^*}^I \bar{x}_{i^*}$.

Now $\sum_i \sum_{k=1}^{K_i} \xi_i^{(*,k)}(x_i^k - \overline{x}_i) \leq 0$ implies that we can again as in Case 1 set $\theta_i^k = \xi_i^{(*,k)}$, assume without loss that $\sum_k \theta_i^k \leq 1$, and define $\theta_i^0 = 1 - \sum_k \theta_i^k$. Then we may set $z_i^0 = \bar{x}_i$ when $\theta_i^0 > 0$ and $z_i^k = x_i^k$ when $\theta_i^k > 0$ and then we have (ignoring terms where $\theta_i^k = 0$)

$$\sum_i \sum_{k=0}^{K_i} \theta_i^k z_i^k \leq \sum_i \bar{x}_i$$

so that if we define an allocation by $y_i = \sum_{k=0}^{K_i} \theta_i^k z_i^k$, and recall that $x_{i^*}^k \succ_{i^*}^I \bar{x}_{i^*}$, we conclude that the allocation (y_i) empirically dominates $(\bar{x}_i)$.

5.2 Proof of Theorem 2

For this proof we start by constructing the same matrix as in the proof of Theorem 3 but with $N = 1$, and where we now add a row $\mathbf{1}_* - \mathbf{1}_k$ for each k to capture the inequality $u^k \leq \bar{u}$. The idea is to consider the same collection of linear inequalities as before, but where we in addition require that the level of utility in the new observation exceeds that of any existing observation in the data. Consider a solution to the dual. Again when restricted to the incidence matrix there is a collection of oriented cycles in the graph, each cycle being associated with a weight, and the total weight ascribed to an incidence vector is the sum of all weights associated to cycles in which the incidence vector appears. A cycle includes no repetitions of nodes.

Because the individual dataset $\{(p_i^k, x_i^k)\}_{k=1}^{K_i}$ is rationalizable, we may assume without loss of generality that every such cycle involves an edge of the type connecting (i,k) to $(i,*)$. This is because otherwise, along all elements of the cycle, rationalizability implies that $p_i^{k_j} \cdot (x_i^{k_{j+1}} - x_i^{k_j}) = 0$, and thus the weighted sum of vectors across that cycle is zero. Removing them does not affect the total weighted sum of rows.

By the same argument as in Theorem 3, if $\mathcal{C}$ denotes the set of cycles, each of them with weight $\mu(c)$, we know that a cycle has an edge connecting (say) (k) to $(*)$, where $\xi^{(k,*)} > 0$ and that in consequence $x^k \succeq^I \bar{x}$. What is different from the proof of Theorem 3 is that now the cycle may involve an edge going from (say) (l) to $(*)$ which was added from a row $\mathbf{1}_* - \mathbf{1}_l$ due to the inequality $u^l \leq \bar{u}$.

Now as before there are two cases to contend with. First, when $\xi^M > 0$ we obtain as before that $\sum_k \xi^{(k,*)}(x^k - \bar{x}) \ll 0$. This means that there is a convex combination $\theta^- \bar{x} + \sum_k \theta^k x^k \ll \bar{x}$ with support in $\bar{x}$ and the $x^k \succeq^I \bar{x}$ (as $\theta^k = \xi^{(k,*)} > 0$ means that the argument in previous paragraph applies). Second, when $\xi^M = 0$ then we must have $\xi^k > 0$ for some k. This may again lead to the same case as in Theorem 3, or it may be the case that $\xi^{(k,*)} = 0$ for all k and we have a strict cycle involving the new $\bar{x} \succeq^R x^l$ edges. This would be a violation of GARP.

5.3 Proof of Theorem 1

The starting point for proving this theorem is the system of linear inequalities introduced by Varian (1982) for this problem. Indeed, by Varian's Fact 4 (Varian (1982)), $\bar{y}$ is revealed worse than $\bar{x}$ if and only if there is no solution $q > 0$ to the system of linear inequalities comprised by:

1. $q \cdot \bar{x} \leq q \cdot x^k$ for all k with $x^k \succeq^I \bar{x}$
2. $q \cdot \bar{x} \leq q \cdot x^k$ for all k with $x^k \succeq^I \bar{y}$

3. $q \cdot \bar{x} < q \cdot x^k$ for all k with $x^k \succ^I \bar{x}$

4. $q \cdot \bar{x} < q \cdot x^k$ for all k with $x^k \succ^I \bar{y}$

Set up a matrix to capture this system, with one row for each of the inequalities that are collected in 1-4 above. These rows are of the form $x^k - \bar{x}$. We want $q > 0$ so there is also one row for each $q_h \geq 0$ inequality, and one row for the inequality that $\sum_h q_h > 0$. Consider a dual solution with weights $\theta^k \geq 0$ for each of the inequalities involving $\bar{x}$, and $\eta^k \geq 0$ for the inequalities that involve $\bar{y}$. We use a prime to distinguish revealed preference from strict revealed preference. Let $\xi^h \geq 0$ be the dual variable for the $q_h \geq 0$ inequalities and $\xi^M \geq 0$ for the last $\sum_h q_h > 0$ inequality. The dual then says, for each h,

$$\sum_{\{k:x^k \succeq^I \bar{x}\}} \theta^k (x^k_h - \bar{x}_h) + \sum_{\{k:x^k \succ^I \bar{x}\}} \theta'^k (x^k_h - \bar{x}_h) + \sum_{\{k:x^k \succeq^I \bar{y}\}} \eta^k (x^k_h - \bar{x}_h) + \sum_{\{k:x^k \succ^I \bar{y}\}} \eta'^k (x^k_h - \bar{x}_h) + \xi^h + \xi^M = 0$$

In an abuse of notation, we shall not distinguish between variables with and without prime. Suppose first that $\xi^M > 0$. Then we get that $\sum_k (\theta^k + \eta^k) x^k \ll \bar{x} \sum_k (\theta^k + \eta^k)$, which means that $\sum_k (\theta^k + \eta^k) > 0$ and that we may normalize so that $\sum_k \theta^k + \eta^k = 1$. Set $z^{k^*} \gg x^{k^*}$ for some $\theta^{k^*} + \eta^{k^*} > 0$, and $z^k = x^k$ for all other $k \neq k^*$, so that $\bar{x} = \sum_k (\theta^k + \eta^k) z^k$ with $z^k \succeq^I \bar{x}$ or $z^k \succeq^I \bar{y}$ for each k, and where the comparison becomes $\succ^I$ for $k = k^*$. Notice that we can choose k^* so that $\eta^{k^*} > 0$ because if all the η variables were zero we would have a certificate for the inequalities in 1 and 3 being infeasible; we know, however, that these are feasible.[12]

If instead $\xi^M = 0$ then we must have $\theta^k + \eta^k > 0$ for some k with either $x^k \succ^I \bar{x}$ or $x^k \succ^I \bar{y}$. Again this allows us to assume that $\sum_k \theta^k + \eta^k = 1$ and we get that $\sum_k (\theta^k + \eta^k) x^k \leq \bar{x}$.

[12]Indeed, if we consider only the inequalities and 1 and 3, and if the dataset is rationalizable, then we may choose $q > 0$ to support a rationalizing utility at $\bar{x}$. The resulting dataset, adding the observation $(q, \bar{x})$, must be rationalizable.

5.4 Proof of Theorem 5

We shall omit some details as all these proofs involve similar ideas. Set up the problem as in Theorem 3. The same system of Afriat inequalities for the observed choices, and the unknown price q that supports the new allocation $(\bar{x}_i)$. Now, however, we add inequalities to capture that $\bar{x}_i$ must be affordable at the income that agents derive from selling their endowment at equilibrium prices. In fact impose the inequality $q \cdot (\omega_i - \bar{x}_i) \geq 0$. Let α_i be the dual variable associated to this inequality. Since $\bar{x}_i$ is an allocation of ω_i these will ensure that the inequality holds with equality for all agents. Now we obtain, reasoning as before, that a dual solution implies

$$\sum_i \sum_k \theta_i^k (x_i^k - \bar{x}_i) + \sum_i \alpha_i(\omega_i - \bar{x}_i) + \sum_m \xi^m \mathbf{1}_m + \xi^M \mathbf{1} = 0$$

Suppose first that $\xi^M > 0$ and normalize so that $\sum_k \theta_i^k + \alpha_i \leq 1$. Let $\bar{y}_i = \sum_k \theta_i^k x_i^k + \alpha_i \omega_i + (1 - \sum_k \theta_i^k - \alpha_i)\bar{x}_i$. Then we obtain

$$\sum_i \bar{y}_i \ll \sum_i (1 - \alpha_i)\bar{x}_i \leq \sum_i \bar{x}_i.$$

And as in the previous proof, when $\xi^M = 0$ then one of the strict revealed preference comparisons must get strictly positive weight.

5.5 Proof of Theorem 6.

Normalize the data so that income in each observation equal 1, so we have $I_i^k = 1$ for all k and i. Define the revealed preference relation as before, but now add the comparisons $0 \succeq_i^R k$ when $\bar{p} \cdot x_i^k \leq \bar{p} \cdot \omega_i$ and $0 \succ_i^R k$ when $\bar{p} \cdot x_i^k < \bar{p} \cdot \omega_i$. Then we abuse notation by denoting by $\succeq_i^R$ and $\succ_i^R$ the resulting transitive closures.

Consider a linear system with the following inequalities:

1. $p^k \cdot \bar{x}_i \geq 1$ for all i and k with $0 \succeq_i^R k$.

2. $p^k \cdot \bar{x}_i > 1$ for all i and k with $0 \succ_i^R k$.

3. $\bar{p} \cdot \bar{x}_i \geq \bar{p} \cdot \omega_i$ for all i.

4. $\sum_i \bar{x}_i = \sum_i \omega_i = \bar{\omega}$ (market clearing).

5. $\bar{x}_i \geq 0$.

Set this up as a homogenous system with $NM + 1$ columns: the first M correspond to the unknowns $\bar{x}_{i,m}$ for $i \in N$ and $1 \leq m \leq M$. The last column is used for a normalization variable that will be required to be strictly positive, and then normalized to 1 in any solution. The rows of this matrix correspond to the 5 categories of inequalities in the system. So the last column has -1 for the first two collection of rows, $-I_i =$ for the second collection of rows, where $I_i = \bar{p} \cdot \omega_i$, $-\bar{\omega}_m$ for the following set of rows; then 0 for the non-negative inequality, and finally 1 for the last added row. Let π be the dual variable for the last "normalization" inequality.

$$
\begin{array}{r}
\\
0 \succeq_i^R k \\
\vdots \\
0 \succ_i^R l \\
\vdots \\
i \\
\vdots \\
m \\
\vdots \\
(i,m) \\
\\
\end{array}
\begin{array}{c}
\begin{array}{ccccc} (1,1) & \cdots & (i,m) & \cdots & (N,M) \end{array} \\
\left[\begin{array}{ccccc|c}
0 & \cdots & p_{i,m}^k & \cdots & 0 & -1 \\
\vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & p_{i,m}^l & \cdots & 0 & -1 \\
\vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & p_{i,m}^k & \cdots & 0 & -I_i \\
\vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & 1 & \cdots & 0 & -\bar{\omega}_m \\
\vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & 1 & \cdots & 0 & 0 \\
0 & \cdots & 0 & \cdots & 0 & 1
\end{array}\right]
\end{array}
$$

Let the dual variables be θ_i^k for the first two collection of inequalities, α_i for the next set of inequalities, η^m for the market-clearing inequalities, ξ_i^m for the non-negativity constraint, and π for the very last "normalization" inequality. Now the dual system is

$$\sum_k \theta_i^k p^k + \alpha_i \bar{p} + \eta + \xi_i = 0 \text{ for all } i,$$

and

$$-\sum_i \sum_k \theta_i^k - \sum_i \alpha_i I_i - \eta \cdot \bar{\omega} + \pi = 0$$

Clearly the primal system has a solution if the last inequality is ignored, so we must have $\pi > 0$ in any dual solution. The first system implies that $\eta \leq 0$, so the last system implies that $\sum_{i,k} \theta_i^k + \sum_i \alpha_i > 0$. Define $\beta = -\eta$ and normalize the dual variables so that $\sum_k \theta_i^k + \alpha_i < 1$ for all i. Then we have that

$$\sum_k \theta_i^k p^k + \alpha_i \bar{p} + (1 - \sum_k \theta_i^k - \alpha_i)\xi_i' = \beta \text{ for all } i,$$

as well as

$$\sum_i \sum_k \theta_i^k \sum_i \alpha_i I_i = \beta \cdot \bar{\omega} + \pi.$$

This means that there is a probability measure μ_i for each i on

$$\{k : \bar{p} \succeq_i^R x_i^k\} \cup \{a, b\} \text{ such that } \mathbf{E}_{\mu_i} \tilde{p} = \beta,$$

where $\tilde{p}$ equals p^k on k, $\bar{p}$ on a and ξ_i' on b. And

$$\sum_i \mathbf{E}_{\mu_i} \tilde{I}_i < \beta \cdot \bar{\omega},$$

where $\tilde{I}_i$ is 1 on k, I_i on a and 0 on b.

5.6 Proof of Theorem 7

It is obvious that the existence of these utilities imply that the datasets are rationalizable. We prove the opposite direction.

Let agent i be the consumer i^* in the hypothesis of the theorem. First we argue that the union $D_i \cup D_a$ is rationalizable. Indeed each of the datasets D_i and D_a is rationalizable, so any revealed preference cycle would have to involve an edge $p \cdot x \geq p \cdot x'$ for $(p, x) \in D_i$ and $(p', x') \in D_a$. This is, however, not possible as $x < x'$ by definition of the consumer i.

Now let u be a rationalization of $D_i \cup D_a$ and define $u_i = v = u$. By

Afriat's theorem, we may take these utilities to be increasing and concave. Let u_j, for $j \neq i$ be an arbitrary rationalization of D_j. For any observed price p^k, the observed allocation (x_i^k) and these utilities satisfy the property in the statement of the theorem. For any unobserved price p, let $x \in \text{argmax}\{v(z) : z \in \mathbf{R}_+^m \text{ and } p \cdot z \leq 1\}$ and choose $x_i = x$ and $x_j = 0$ for $j \neq i$. Since $u_i = v$ the resulting allocation satisfies the statement in the theorem.

References

Abdulkadiroğlu, A. and T. Sönmez (2003): "Ordinal efficiency and dominated sets of assignments," *Journal of Economic Theory*, 112, 157–172.

Afriat, S. N. (1967): "The construction of utility functions from expenditure data," *International economic review*, 8, 67–77.

Allen, R. and J. Rehbeck (2020a): "Counterfactual and welfare analysis with an approximate model," *arXiv preprint arXiv:2009.03379*.

——— (2020b): "Satisficing, aggregation, and quasilinear utility," *Available at SSRN 3180302*.

Antonelli, G. (1886): "Sulla Theoria Matematica Della Economia Politica," Privately published.

Aziz, H., F. Brandl, and F. Brandt (2015): "Universal Pareto dominance and welfare for plausible utility functions," *Journal of Mathematical Economics*, 60, 123–133.

Bachmann, R. (2004): "Rationalizing allocation data: a nonparametric Walrasian theory when prices are absent or non-Walrasian," *Journal of Mathematical Economics*, 40, 271–295.

——— (2006a): "Testable implications of coalitional rationality," *Economics Letters*, 93, 101–105.

——— (2006b): "Testable Implications of Pareto Efficiency and Individual rationality," *Economic Theory*, 29, 489–504.

Berge, C. (2001): *The theory of graphs*, Dover Publications.

Bewley, T. F. (2002): "Knightian decision theory. Part I," *Decisions in economics and finance*, 25, 79–110.

Bewley, T. F. et al. (1987): "Knightian decision theory, Part II: Intertemporal problems," *Cowles Foundation Discussion Papers*, 835.

Blundell, R., M. Browning, and I. Crawford (2007): “Improving revealed preference bounds on demand responses,” *International Economic Review*, 48, 1227–1244.

——— (2008): “Best nonparametric bounds on demand responses,” *Econometrica*, 76, 1227–1262.

Blundell, R. W., M. Browning, I. Crawford, B. De Rock, F. Vermeulen, and L. Cherchye (2015): “Sharp for SARP: Nonparametric bounds on the behavioural and welfare effects of price changes,” *AEJ Microeconomics*, 7.

Bogomolnaia, A. and E. J. Heo (2012): “Probabilistic assignment of objects: Characterizing the serial rule,” *Journal of Economic Theory*, 147, 2072–2082.

Bogomolnaia, A. and H. Moulin (2001): “A new solution to the random assignment problem,” *Journal of Economic theory*, 100, 295–328.

Border, K. C. (2015): “Miscellaneous Notes on Optimization Theory and Related Topics,” Mimeo, Caltech.

——— (2020): “Alternative Linear Inequalities,” `http://www.its.caltech.edu/~kcborder/Notes/Alternative.pdf`, accessed: 2020-1-25.

Bossert, W. and Y. Sprumont (2002): “Core rationalizability in two-agent exchange economies,” *Economic theory*, 20, 777–791.

Brown, D. J. and C. Calsamiglia (2007): “The nonparametric approach to applied welfare analysis,” *Economic Theory*, 31, 183–188.

Brown, D. J. and F. Kubler (2008): “Refutable Theories of Value,” in *Computational Aspects of General Equilibrium Theory*, Springer, 1–10.

Brown, D. J. and R. L. Matzkin (1996): “Testable restrictions on the equilibrium manifold,” *Econometrica*, 64, 1249–1262.

Brown, D. J. and C. Shannon (2000): "Uniqueness, stability, and comparative statics in rationalizable Walrasian markets," *Econometrica*, 68, 1529–1539.

Carroll, G. (2010): "An efficiency theorem for incompletely known preferences," *Journal of Economic Theory*, 145, 2463–2470.

Carvajal, A. (2004): "Testable restrictions on the equilibrium manifold under random preferences," *Journal of Mathematical Economics*, 40, 121–143.

——— (2010): "The testable implications of competitive equilibrium in economies with externalities," *Economic theory*, 45, 349–378.

Carvajal, A., I. Ray, and S. Snyder (2004): "Equilibrium behavior in markets and games: testable restrictions and identification," *Journal of Mathematical Economics*, 40, 1–40.

Carvajal, A. and X. Song (2018): "Testing Pareto efficiency and competitive equilibrium in economies with public goods," *Journal of Mathematical Economics*, 75, 19–30.

Chambers, C. P. and F. Echenique (2009): "Supermodularity and preferences," *Journal of Economic Theory*, 144, 1004–1014.

——— (2016): *Revealed preference theory*, vol. 56, Cambridge University Press.

Cherchye, L., I. Crawford, B. De Rock, and F. Vermeulen (2016): "Gorman revisited: nonparametric conditions for exact linear aggregation," *SERIEs*, 7, 203–220.

Cherchye, L., B. De Rock, and F. Vermeulen (2009): "Opening the black box of intrahousehold decision making: Theory and nonparametric empirical tests of general collective consumption models," *Journal of Political Economy*, 117, 1074–1104.

Cherchye, L., T. Demuynck, and B. De Rock (2011): “Testable implications of general equilibrium models: an integer programming approach,” *Journal of Mathematical Economics*, 47, 564–575.

Doğan, B. and K. Yildiz (2016): “Efficiency and stability of probabilistic assignments in marriage problems,” *Games and Economic Behavior*, 95, 47–58.

Dow, J. and S. R. da Costa Werlang (1988): “The consistency of welfare judgments with a representative consumer,” *Journal of Economic Theory*, 44, 269–280.

Dubra, J., F. Maccheroni, and E. A. Ok (2004): “Expected utility theory without the completeness axiom,” *Journal of Economic Theory*, 115, 118–133.

Eliaz, K. and E. A. Ok (2006): “Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences,” *Games and Economic Behavior*, 56, 61–86.

Fon, V. and Y. Otani (1979): “Classical welfare theorems with non-transitive and non-complete preferences,” *Journal of Economic Theory*, 20, 409–418.

Gale, D. and A. Mas-Colell (1975): “An equilibrium existence theorem for a general model without ordered preferences,” *Journal of Mathematical Economics*, 2, 9–15.

——— (1977): “On the role of complete, transitive preferences in equilibrium theory,” in *Equilibrium and disequilibrium in economic theory*, Springer, 7–14.

Gorman, W. M. (1953): “Community preference fields,” *Econometrica*, 21, 63–80.

Graaff, J. d. V. (1967): *Theoretical welfare economics*, Cambridge University Press.

Hashimoto, T., D. Hirata, O. Kesten, M. Kurino, and M. U. Ünver (2014): "Two axiomatic approaches to the probabilistic serial mechanism," *Theoretical Economics*, 9, 253–277.

Hicks, J. R. (1939): "The foundations of welfare economics," *The economic journal*, 49, 696–712.

Kaldor, N. (1939): "Welfare propositions of economics and interpersonal comparisons of utility," *The Economic Journal*, 49, 549–552.

Kubler, F. (2003): "Observable restrictions of general equilibrium models with financial markets," *Journal of Economic Theory*, 110, 137–153.

Manea, M. (2008): "A constructive proof of the ordinal efficiency welfare theorem," *Journal of Economic Theory*, 141, 276–281.

McLennan, A. (2002): "Ordinal efficiency and the polyhedral separating hyperplane theorem," *Journal of Economic Theory*, 105, 435–449.

Nishimura, H., E. A. Ok, and J. K.-H. Quah (2017): "A comprehensive approach to revealed preference theory," *American Economic Review*, 107, 1239–63.

Ok, E. A. (2002): "Utility representation of an incomplete preference relation," *Journal of Economic Theory*, 104, 429–449.

Rigotti, L. and C. Shannon (2005): "Uncertainty and risk in financial markets," *Econometrica*, 73, 203–243.

Samuelson, P. A. (1956): "Social indifference curves," *The Quarterly Journal of Economics*, 70, 1–22.

Shafer, W. and H. Sonnenschein (1975): "Equilibrium in abstract economies without ordered preferences," *Journal of Mathematical Economics*, 2, 345–348.

——— (1982): "Market demand and excess demand functions," *Handbook of mathematical economics*, 2, 671–693.

Varian, H. R. (1982): "The nonparametric approach to demand analysis," *Econometrica*, 50, 945–973.

Weymark, J. A. (1985): "Remarks on the first welfare theorem with nonordered preferences," *Journal of Economic Theory*, 36, 156–159.