ON THE CONSISTENCY OF DATA WITH BARGAINING THEORIES

CHRISTOPHER P. CHAMBERS AND FEDERICO ECHENIQUE

Abstract. We develop observable restrictions of well-known theories of bargaining over money. We assume given a finite data set of bargaining outcomes, including data on allocations and outside options, but no information on utility functions. We ask when a given theory could have generated the data. If the outside options are fixed and symmetric, we show that Nash, utilitarian, and the egalitarian max-min bargaining solutions are all observationally equivalent. These theories are in turn characterized by a simple test of comonotonicity of bargaining outcomes.

If the outside options are arbitrary, we establish different testing procedures for each of the theories under consideration. The main application of our result is to testing the tax code for compliance with the principle of equal loss.

1. Introduction

This paper is an inquiry into the testable implications of bargaining theory. We have in mind the allocation of a single-dimensional resource: we can essentially focus on the allocation of money amongst a set of agents. We suppose that we have available certain data on how money was divided amongst a fixed number of agents. The data describe final allocations of money, and the agents' outside options, but include no information on agents' preferences, or on the method (or protocol, or

---

Chambers is affiliated with the Department of Economics, University of California, San Diego. Echenique is affiliated with the Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena CA 91125. Emails: cpchambers@ucsd.edu (Chambers), fede@caltech.edu (Echenique). We are grateful to Thomas Demuynck, Bram De Rock, Diego Dominguez, Kfir Eliaz, Ivana Komunjer, Cesar Martinelli, and William Thomson for comments and suggestions. We want to especially thank the coeditor Martin Osborne and three anonymous referees, for many improvements on the first draft we submitted.

extensive-form game) that led to the division. We want to know when observed allocations are consistent with standard bargaining theory.

There are several well-known theories that could explain a given collection of data. We focus on the *utilitarian*, *Nash*, and *egalitarian* models. The utilitarian model divides the resource so as to maximize the sum of agents' utilities. The Nash bargaining model (after Nash (1950)) seeks to maximize the product of the gains from splitting the resource. Finally, the egalitarian solution (commonly identified with Rawls (1971)) chooses a division to maximize the utility of the worst off agent. Our goal here is to investigate the implications of these theories for certain kinds of data sets.

Data sets consist of observations of how a resource is divided amongst a set of agents, and of the relevant "disagreement point." That is, we observe how much money was at stake, how it was divided among the agents, and which outcome would have prevailed if the agents had failed to agree on a division. We do not, however, have any data on agents' preferences, nor on the protocol (or underlying strategic interactions) that produced the observed allocations.

Such data are used by Hamermesh (1973), for example. He uses data on union wage bargaining, where, if negotiations break down, the disagreement outcome for all individuals is known to be zero. More generally, the data could be the outcome of bankruptcy liquidation proceedings, or government subsidies. We investigate the restrictions that each of the three models, utilitarian, Nash, and egalitarian, place on the allocations of money. Essentially, we want to test the hypotheses that resources are allocated according to these bargaining theories when utility functions are assumed to be monotone and concave, but otherwise can differ across individuals.

We present two sets of results.

First, in Section 4, we assume that the disagreement points are fixed across observations, and the same for all agents. A case in point is the wage bargaining data mentioned above. *We discover that the empirical content of the three models is identical: No data set of the kind we* *have assumed will ever allow us to distinguish between them.* A data set either refutes all three or is consistent with all three.

Furthermore, the theories have very weak predictive power. The only empirical prediction of any of these theories is that data should be perfectly strictly ordinally correlated, or comonotonic, and that each agent should get at least a positive amount of consumption if any other agent does. This means that when the total amount of resource increases, all agents must receive a higher amount of resource.

In fact, we show more. We show that with our assumption on data sets, comonotonicity characterizes the empirical content of a large class of social welfare functions, namely, any which takes a *generalized utilitarian* form: that is $\sum_{i\in N} g(u_i(x_i))$. The utilitarian and Nash models are special cases. Our result implies a rather strong form of observational equivalence. Whenever the data are consistent with one of these theories (say Nash bargaining), then *the same* rationalizing utility functions serve to rationalize the observed data as generated by any of the other theories. As far as we know, our result is the first to document this strong form of observational equivalence in the revealed preference literature. We are considering three "non-nested" models, and given a rationalizable data set, we can find utility functions that serve to rationalize the data using any of the three models.

Secondly, in Section 5, we turn to data in which the disagreement point can vary. We observe here that the result from Section 5 readily extends to the utilitarian model, and we present a simple result on the testable implications of the Nash bargaining solution. The most interesting case, though, deals with the egalitarian model of bargaining.

Under the additional assumption that utility functions are the same, the egalitarian model embodies the principle of "equal gain." The increase in utility between the final outcome and the disagreement point should be equal across agents.

The most interesting empirical application is to taxation. We can interpret data on taxation (specifically, the tax code) as bargaining data: the disagreement point represents agents' post-tax incomes, and the division of money represents the amount they earn before taxes.

Data of this kind are readily available from the tax code, and do not require collecting data on agents' incomes.

The egalitarian theory is then a classical principle of taxation. Instead of equal gains we require "equal loss" (Young, 1988). The principle of equal loss has a long history in the economics of taxation.[1] We are concerned with the empirical problem of testing a tax code for compliance with the principle of equal loss. We present a test of this theory in the case when utility is unknown. Young (1990) studies the same problem, but using a parametric estimation approach to find the best-fitting utility index to tax data in the United States. We present instead a non-parametric test, which can be applied to the data used by Young.

Section 5 discusses the case when utilities may differ among agents, and presents an extension to when utility is required to be concave, as it is in the rest of the paper. The results of Section 5 have an interesting byproduct: the testable implications of Hotelling's model of spatial competition (Hotelling, 1929). Subsection 5.8 demonstrates how this problem is a special case of the environment studied in section 5.

1.1. **Related literature.** The closest papers to ours may be Cherchye et al. (2011) and Carvajal and González (2011). These are independently conducted investigations into the testable implications of Nash bargaining.

Cherchye, Demuynck, and De Rock consider a model where a pair of agents bargain over consumption decisions; so the framework is different from our focus on bargaining over money. They assume that the disagreement points vary endogenously because individual agents have the option of making consumption purchases on their own, and they characterize the rationalizable data as those that satisfy a system of quadratic inequalities. Recognizing that such a system is hard to solve, they provide a sufficient condition, and a necessary condition,

[1] As Young (1990) notes, it was championed by John Stuart Mill, and spawned a large literature on the normative virtues of equal sacrifice.

which can be operationalized computationally. Finally, they carry out a laboratory experiment and show how one can use their tests.

Carvajal and González suppose that consumption is over monetary units (as we do), and develop polynomial tests of the Nash bargaining model under various hypotheses about the behavior of the disagreement point. Most of their tests characterize rationalizable data as those that satisfy a system of quadratic inequalities. They use the Tarski-Seidenberg algorithm to construct direct tests of rationalizability in terms of data alone.

The setup and methodology in both Cherchye et al. (2011) and Carvajal and González (2011) are distinct from ours, and perhaps closest to our discussion in Section 5.2. In fact, probably the method we suggest there can be applied in their frameworks, and vice versa. The boundary of the problems we can solve in revealed preference analysis is given by polynomial problems, such as the ones they analyze. It is interesting to see complementary approaches emerge.

The recent contribution of Chiappori et al. (2012) investigates the empirical content of Nash bargaining. There are several important differences between that work and ours. The main difference is that their framework assumes disagreement points are unobserved. Instead, they suppose that some vector of underlying, observable Euclidean characteristics uniquely determine both the utility functions of agents, as well as the disagreement point. Without assuming any kind of structure on the joint dependence of disagreement point and utility on these underlying characteristics, their model obviously has no testable implications (this is their Proposition 2). To have any empirical content, they must assume some structure on the dependence of the utility function and disagreement point on these characteristics. They assume that this dependence is known to satisfy certain properties (differentiability and "exclusion restrictions") both within and across characteristics. By contrast, in our model, disagreement point observations are part of the observed data, and this leads to the falsifiability of the model.

The other main distinction between their work and ours is that they are concerned with understanding the testable implications of the model in a continuous sense–the implications of the model if we could observe the division across *all* possible problems. Our work, on the other hand, assumes only that a finite number of possible division problems are observed (with their solutions). The distinction in the two approaches can be best understood by considering the classical demand model: their approach is analogous to characterizing rationalizability by conditions on the Slutsky matrix, while our approach is analogous to Afriat's (Afriat, 1967) discussion of finite data sets which are rationalizable. Their work also notes that the testable implications of the Nash, utilitarian, and maxmin model are identical in certain frameworks.

De Clippel and Eliaz (2011) also provide an interesting study of the empirical content of a particular bargaining solution, which they call the *fallback* solution (which shares some ideas with the maxmin solution). Their framework is a general ("Arrovian") choice environment, where two agents decide from a finite choice set. The paper by Kıbrıs (2011) also studies bargaining solutions (albeit in a claims framework) and derives a revealed preference relation from the choices made by the solutions. Kıbrıs finds conditions on the solutions so that the implied revealed preference relations is rational.

Earlier works discussing the empirical content of Nash bargaining, usually assuming all individuals are identical and risk neutral, include Hamermesh (1973) and Bowlby and Schriver (1978). Svejnar (1980) provides a critique of these ideas.

As earlier noted, Young (1990) constructs a test of the maxmin hypothesis, using empirical data on US income taxes from 1957-1987. His approach is estimation-based, and he finds that tax data are reasonably close to predicted data from the maxmin model in most years (there are exceptions). He assumes specific parametric forms for the utility function. By contrast, we provide an exact test of the maxmin model, assuming no parametric functional form. Young (1988) provides a kind of exact empirical test of the maxmin model, assuming the solutions to all possible problems are observed, and further assuming observations *across* different populations.

2. The Theories

We consider the most commonly used (cooperative) theories of bargaining. We assume an environment where some quantity $m \in \mathbf{R}$ of a single-dimensional resource (e.g money) is available, and a group of n agents needs to decide on an allocation of m. Each agent i has an outside option that they can turn to instead of agreeing to a bargaining outcome; the outside option is represented by a number d_i. The vector $d = (d_i)$ is a *disagreement point.* The set

$$B(m,d) = \{(x_1, \dots, x_n) \in \mathbf{R}^n_+ : \sum_{i=1}^{n} x_i \leq m \text{ and, for all } i, x_i \geq d_i\}$$

represents all the allocations of m in which each agent gets at least their outside option. The disagreement point is observed and known.

A bargaining theory uses information on agents' preference to predict an outcome in $B(m,d)$. Suppose that each agent i is described by a strictly monotonic and concave utility function $u_i : \mathbf{R}_+ \to \mathbf{R}$.

The *utilitarian* theory says that m should be divided so as to maximize the sum $\sum_{i=1}^{n} u_i(x_i)$ over all $x = (x_1, \dots, x_n) \in B(m,d)$. We shall consider a generalization of the utilitarian theory, where for some function $g : A \subseteq \mathbf{R} \to \mathbf{R}$, the sum $\sum_{i=1}^{n} g(u_i(x_i) - u_i(d_i))$ is maximized over $B(m,d)$.

The *Nash bargaining* theory predicts a choice in $B(m,d)$ that maximizes the so-called *Nash product,*

$$\prod_{i=1}^{n} \left[u_i(x_i) - u_i(d_i)\right].$$

Note that the Nash bargaining theory is a special case of our generalization of the utilitarian theory, letting $g = \log$.

Finally, the *egalitarian* (or maxmin) theory says that $x \in B(m,d)$ should be chosen to maximize

$$\min_{i \in N} \left[u_i(x_i) - u_i(d_i)\right]$$

These three theories have both positive and normative interpretations. Under the positive interpretation, it is evident that we may want to understand the empirical content of the theory. The theories are commonly assumed to predict an outcome in applied economic models. Probably the most commonly used is the utilitarian model, but the Nash solution also finds applications. From macroeconomics to contract theory and applied mechanism design, authors often use the Nash solution as a "reduced form" model to capture the outcome of some bargaining stage.

From the normative viewpoint, our three theories have know axiomatizations which relate them to basic principles of justice or social welfare (see, for example, Thomson (1981), Thomson (2010), Lensberg (1987), Thomson and Lensberg (1989), Kalai (1977)). If empirical outcomes are incompatible with the utilitarian rule, for example, one can advise the agents to change their behavior in order to comply with some attractive principles of justice or social welfare. This empirical "spin" on a normative theory is similar to Savage's famous reaction to exhibiting the Ellsberg paradox, and to the use of empirical data in applied market design.

3. The Data

We assume a finite collection of observations of bargaining outcomes. A *data set* is a set $D = \left\{\left(d^k, x^k\right)\right\}_{k=1}^{K}$. Each observation k specifies a pair $(d^k, x^k) \in \mathbf{R}^{2n}$, where x^k is an allocation in $B(\sum_{i \in N} x_i^k, d^k)$. The interpretation of D is that we observe the outcome of bargaining for different levels of resource available, and different disagreement points.

Any study of the empirical content of a theory depends crucially on what one assumes is observable. If we observe m, d, and agents' utility functions, then the theories we described in Section 2 are all testable, and each one of them make predictions that are not made by the others: so they can be distinguished empirically when one has available data on all the primitives of the theory.

In contrast, we assume that utility functions are not observable. Our assumption follows the mainstream revealed preference view of preference and utility: utilities are purely theoretical constructs and do not have any meaning beyond what they predict about agents' behavior–in this case joint behavior. Utilities are not observable, or even meant to be observable.

The revealed preference view is rooted in the use of field data in economics. Most data sets in economics consist of field data, and they do not normally contain information on utility functions. In the particular case of bargaining theory, the econometric studies of bargaining use data with no information on utility functions (see, for example, Hamermesh (1973)).

Experimental data can, and often do, collect partial information on utility functions by the design of the experiment or by using supplementary surveys. Many experimentalists are, however, skeptical about the idea that one can control utility effectively in the laboratory. In experiments specifically designed to test bargaining theory, attempts have been made to assume as little as possible about agents' preferences. Al Roth, in a survey of the experimental literature on bargaining (see pages 41-43 in Roth (1995)), argues that experiments that assume a specific utility function are problematic, and describes designs which attempt to assume as little as possible about subjects' preferences (typically only assume that they are expected utility maximizers, but nothing about the form of the utility function). Our assumptions about data are obviously in line with such experimental designs.

4. Fixed and symmetric disagreement point

In this section, we suppose that the disagreement point is fixed and symmetric. Our result is that the only aspect of bargaining theory that can be tested is a basic solidarity principle (see Sprumont (2008)).

The punchline is twofold. First, the three models of Section 2 have identical testable implications. Thus, for the kinds of data described in Section 3, if the disagreement point is fixed and symmetric, then the three most popular models of cooperative bargaining are observationally equivalent.

Second, the empirical predictions of these models are relatively weak. If one agent's consumption increases, then so does the consumption of all remaining agents.

With a fixed disagreement point, we assume that the disagreement point is normalized to 0. A data set (Section 3) then takes the special form $\{(x^k)\}_{k=1}^K$, where $x^k \in \mathbb{R}^n_+$. Importantly, the disagreement point must be the same for all observations. This assumption can be interpreted in two ways. First, we can suppose that the disagreement point is observed, and normalize the data accordingly. Second, we can suppose that the disagreement point is unobserved but fixed; we then must choose our disagreement point, and our test is a joint test of our particular bargaining solution together with the chosen disagreement point.

The data used by Hamermesh (1973) is a case in point. Hamermesh observes bargaining outcomes, but not the disagreement point. It is sensible to assume that, were negotiations to break down, the agents have some fixed outside option to which they can revert.[2]

We shall consider the general model described in Section 2. Let $g : [0, \infty) \to \mathbf{R} \cup \{-\infty\}$ be a strictly increasing, smooth, and concave function. We will say that data $\{(x^k)\}_{k=1}^K$ are *g-rationalizable* if there exist strictly monotonic, smooth, and strictly concave functions u_i for which $u_i(0) = 0$ and $u_i'(0) = \infty$ (Inada conditions), and for which $\sum_{i\in N} g(u_i(x_i^k)) \geq \sum_{i\in N} g(u_i(y_i))$ for all allocations $(y_1, \ldots, y_n) \in B(\sum_i x_i^k, 0)$ and $k = 1, \ldots, K$. As we remarked in Section 2, the utilitarian and Nash model are special cases of g-rationalizability.

Finally, data $\{(x^k)\}_{k=1}^K$ are *maxmin rationalizable* if there exist strictly monotonic and strictly concave u_i, normalized so that $u_i(0) = 0$, for which $\min_{i\in N} u_i(x_i^k) \geq \min_{i\in N} u_i(y_i)$ for all $(y_1, \ldots, y_n) \in B(\sum_i x_i^k, 0)$ and $k = 1, \ldots, K$.

Our result establishes a statement on the data that is equivalent to rationalizability by these strategies: this statement is a form of nonparametric test for bargaining theory. We say that data $\{x^k\}_{k=1}^K$ are

[2]The assumption may be good "discipline," even when it is unlikely to hold exactly. By choosing an unobserved disagreement point arbitrarily, one may be able to explain any set of observations (see our discussion in Section 1.1). So assuming a fixed disagreement point, when the disagreement outside options are actually unobserved, is probably a sensible research strategy.

comonotonic if for all $i, j \in N$ and all k, l, $x_i^k < x_i^l$ implies $x_j^k < x_j^l$, and for all $i, j \in N$, $x_i^k = 0$ if and only if $x_j^k = 0$. Comonotonicity requires that outcomes are perfectly strictly ordinally correlated (when 0 is also considered an outcome).

Theorem 1. *Given data $\{x^k\}_{k=1}^K$ and a strictly increasing concave g, the following are equivalent.*

(1) *The data are comonotonic.*
(2) *The data are g-rationalizable.*
(3) *The data are maxmin rationalizable.*

Remark 2. Our original proof of this theorem only established the equivalence of the utilitarian model, the Nash model, and the maxmin model. An anonymous referee (at the *American Economic Review*) showed us how to generalize the result to the one stated above.

Remark 3. Dropping the hypothesis of smoothness of the u_i functions and of g results in a weaker notion of comonotonicity. Dropping the hypothesis that u is strictly concave can result in models with no testable implications on our class of data. For example the utilitarian model with linear utility functions is not testable.

Remark 4. We fix g, thereby fixing a theory, and ask if there are utility functions which rationalize the observations. In principle the rationalizing utility functions *could* depend on the particular g under consideration. For example, the utilities that rationalize some data set as coming from Nash bargaining may differ from the utilities which give a utilitarian rationalization. *Surprisingly, it turns out that the utilities we construct in the proof serve to rationalize the data for any g.* This was shown to us by the same anonymous referee.

In fact, it is evident from our proof that the constructed utilities also allow rationalization by any symmetric, monotonic, and quasiconcave (even quasiconcavity can be weakened) function $\varphi : \mathbf{R}^n \to \mathbf{R}$, independently of whether or not φ is additively separable [3]

[3]The converse statement, that any symmetric, monotone, and quasiconcave social welfare function generates comonotonic data, is not true. It seems that the additive welfare functions we consider are the most general ones with this property.

Thus the models we consider are observationally equivalent in a particularly strong sense. Rationalizable data allow us to fix the unobservable utility functions in a way that is consistent with any of the models under consideration. We cannot differentiate one model from another in terms of their implied behavior about rationalizing utility functions.

Remark 5. We could also ask about rationalization by "weighted" versions of the rules. For example, with weights $\alpha \in \mathbf{R}^n_{++}$, a *weighted g rule* might be one which is chosen to maximize $\sum_{i\in N} g(\alpha u_i(x_i))$ across $B(x,0)$, subject to the constraint that $\sum_{i\in N} x_i = x$. By Theorem 1, the only implication of the maximization of such rules is comonotonicity. This can be seen by finding u_i functions which g-rationalize the data, and then rescaling. A similar statement holds for maxmin rationalizability.

Proof. It is clear from the first order conditions that if the data are either g-rationalizable or maxmin rationalizable, then they are comonotonic.[4]

For the other direction, we show something slightly stronger: If the data are comonotonic, then there exist strictly concave, continuous, and increasing u_i such that, if $\varphi : [0,\infty) \to \mathbf{R} \cup \{-\infty\}$ is a monotonic, symmetric and quasiconcave function, then $\varphi(u_1(x^k_1),\dots,u_n(x^k_n)) \geq \varphi(u_1(y_1),\dots,u_n(y_n))$ for all allocations $(y_1,\dots,y_n)$ satisfying $\sum_{i\in N} x^k_i = \sum_{i\in N} y_i$.[5] As a special case, we have $\varphi(z_1,\dots,z_n) = \sum_{i=1}^n g(z_i)$. Note the order of the quantifiers used

[4] The argument for the utilitarian model is: Suppose that all agents' marginal utilities are equalized at $x \in B(m,0)$ and at $x' \in B(m',0)$ with $m = \sum_i x_i$ and $m' = \sum_i x'_i$. Suppose that $m < m'$. Some agent i must have $x'_i > x_i$; then all agents' marginal utilities must be smaller at the allocation x' than at x and the concavity of utility implies that all $x'_i > x_i$ for all i. The argument is almost identical for generalized g-utilitarianism. For the maxmin model, it is clear that all agents must have equal utility at a given allocation. Thus, if one agents' utility goes up, so must all others.

[5] Symmetry means that if σ is a permutation on $\{1,\dots,n\}$ then $\varphi(x_{\sigma(1)},\dots,x_{\sigma(n)}) = \varphi(x_1,\dots,x_n)$. Monotonicity here means that if $x_i > y_i$ for all i, then $\varphi(x_1,\dots,x_n) > \varphi(y_1,\dots,y_n)$.

above: the same profile of utility functions $u_1, \dots, u_n$ works across all φ.

To this end, we suppose the data are comonotonic, and ignore replications as well as points where every agent consumes 0. Without loss of generality, let us suppose that $x_i^1 < x_i^2 < \dots x_i^K$ for all $i \in N$ (that this is possible follows from comonotonicity). Below we construct a profile of utility functions $u_1, \dots, u_n$ with the property that for all $k = 1, \dots, K$, $\sum_{i\in N} u_i(x_i^k)$ is maximal across all allocations $y_1, \dots, y_n$ for which $\sum_{i\in N} x_i^k = \sum_{i\in N} y_i$, and $\min_{i\in N} u_i(x_i^k)$ is also maximal across all such allocations; it follows that, since each u_i is strictly increasing, $u_i(x_i^k) = u_j(x_j^k)$ for all $i, j \in N$.

We first argue that such a construction suffices to establish the result: Let φ be as above, and suppose by way of contradiction that there is a k and a feasible allocation $(y_1, \dots, y_n)$ for which $\varphi(u_1(y_1), \dots, u_n(y_n)) > \varphi(u_1(x_1^k), \dots, u_n(x_n^k))$. Note then, by symmetry of φ, that for any permutation of the agents $\sigma : N \to N$, $\varphi(u_{\sigma(1)}(y_{\sigma(1)}), \dots, u_{\sigma(n)}(y_{\sigma(n)})) = \varphi(u_1(y_1), \dots, u_n(y_n))$. Quasiconcavity of φ then implies that

$$\varphi\left(\sum_{i\in N}\frac{u_i(y_i)}{n}, \dots, \sum_{i\in N}\frac{u_i(y_i)}{n}\right) > \varphi(u_1(x_1^k), \dots, u_n(x_n^k)).$$

By strict monotonicity of φ, and using the fact that $u_i(x_i^k) = u_j(x_j^k)$ for all $i, j \in N$, this implies that $\sum_{i\in N}\frac{u_i(y_i)}{n} > \sum_{i\in N}\frac{u_i(x_i^k)}{n}$, contradicting the that $\sum_{i\in N} u_i(x_i^k) \geq \sum_{i\in N} u_i(y_i)$ for all feasible allocations $y_1, \dots, y_n$.

We finish the proof by constructing, for each i, a strictly decreasing, continuous, and positive function f_i, with the property that if we set u_i to be the integral of f_i, then the profile of utility functions $(u_1, \dots, u_n)$ works as required by the first part of the proof.

We proceed by induction. We ensure that, for each $i \in N$ and each k, the following are true:

(1) $\int_0^{x_i^k} f_i(x)dx = \int_0^{x_j^k} f_j(x)dx$
(2) $f_i(x_i^k) = f_j(x_j^k)$.

Actually, there are many possible constructions of such f_i. We present one. For K, define $f_i(x_i^K) = 1$ for all $i \in N$; for $x > x_i^K$, we define $f_i(x)$ to be any strictly decreasing function, taking value everywhere less than 1 and rendering f_i continuous.

Now, let k be arbitrary, and suppose that $f_i(x)$ has been defined for all $x \geq x_i^k$. We assume that for all $k' \geq k$, $f_i(x_i^{k'}) = f_j(x_j^{k'})$ and

$$\int_{x_i^k}^{x_i^K} f_i(x)dx = \int_{x_j^k}^{x_j^K} f_j(x)dx \text{ for all } i, j \in N.$$

Recall that we have $x_i^1 < x_i^2 < \ldots x_i^K$.

First, in case $k = 1$, we define for each agent j, $f_j(0) = +\infty$. For $k > 1$, we choose a finite $f_j(x_j^{k-1})$ but we must choose it to be sufficiently large. Specifically, let z be large enough so that there is $\varepsilon > 0$ for which $z(x_j^k - x_j^{k-1}) - \varepsilon > \max_{i \in N} f_i(x_i^k)(x_i^k - x_i^{k-1}) + \varepsilon$ for all j. Then we can then set for all j $f_j(x_j^{k-1}) = z$. Observe that, given $f_j(x_j^{k-1})$ and $f_j(x_j^k)$, for any $\varepsilon > 0$ and any

$$y \in \left(f_j(x_j^k)(x_j^k - x_j^{k-1}) + \varepsilon, f_j(x_j^{k-1})(x_j^k - x_j^{k-1}) - \varepsilon\right),$$

we may define f_j continuous and decreasing on $x \in (x_j^{k-1}, x_j^k)$ so that $\int_{x_j^{k-1}}^{x_j^k} f_j(x)dx = y$. This follows as we may approximate the constant functions $h(x) = f_j(x_j^{k-1})(x_j^k - x_j^{k-1})$ and $h^*(x) = f_j(x_j^k)(x_j^k - x_j^{k-1})$ arbitrarily closely pointwise by strictly decreasing and continuous functions. Thus we can complete $f_j(x)$ on $x \in (x_j^{k-1}, x_j^k)$ so that $\int_{x_j^{k-1}}^{x_j^k} f_j(x)dx$ is equalized across all agents (by picking

$$y \in \bigcap_{i \in N} \left(f_i(x_i^k)(x_i^k - x_i^{k-1}) + \varepsilon, f_i(x_i^{k-1})(x_i^k - x_i^{k-1}) - \varepsilon\right)$$

and choosing $f_j(x)$ on $x \in (x_j^{k-1}, x_j^k)$ so that $\int_{x_j^{k-1}}^{x_j^k} f_j(x)dx = y$). In the case of $k = 1$, we must also maintain that $\int_0^{x_j^1} f_j(x)dx < +\infty$.

The functions f_j so constructed satisfy the conditions we ask for: that for all k, $f_j(x_j^k)$ is equalized across j, and $\int_0^{x_j^k} f_j(x)dx$ is equalized across j. By setting $u_j(x) = \int_0^x f_j(x)dx$, we have the required u_j. $\square$

5. Variable disagreement point

We now turn to data as defined in Section 3, where the disagreement point is allowed to vary from one observation to another.

5.1. **The utilitarian model.** The disagreement point constrains the set of possible allocations, $B(m, d)$, but not the objective of maximizing the sum of utilities. In fact, suppose that our data set satisfies $x_i^k > d_i^k$ for all k and i. Then comonotonicity of the x variables is all that is required for utilitarian rationalizability. This is because there exist utility functions which would rationalize the data in the fictitious environment in which each $d_i^k = 0$; it is easily checked that these same utility functions therefore rationalize the given data. Therefore the equivalence between comonotonicity and rationalizability by the utilitarian theory extends to the case of a variable disagreement point.[6]

5.2. **The Nash model.** A data set $D = \left\{\left(d^k, x^k\right)\right\}_{k=1}^K$, with variable disagreement point, is *Nash rationalizable* if there are strictly monotonic, and concave u_i for which

$$\prod_{i \in N}\left[u_i(x_i^k) - u_i(d_i^k)\right] \geq \prod_{i \in N}\left[u_i(y_i^k) - u_i(d_i^k)\right],$$

for all $(y_1, \ldots, y_n) \in B(\sum_{i \in N} x_i^k, d^k)$.

Proposition 6. *A data set D is Nash rationalizable if and only if for all $i \in N$, there are numbers $U_i(d_i^k), U_i(x_i^k), M_i(d_i^k), M_i(x_i^k)$, for $k = 1, \ldots, K$ which solve the following equations: for all $z \in \{x_i^k\}$, $z' \in \{x_i^k, d_i^k\}$, and all i, j and k,*

$$\frac{M_i(x_i^k)}{U_i(x_i^k) - U_i(d_i^k)} = \frac{M_j(x_j^k)}{U_j(x_j^k) - U_j(d_j^k)}$$

and for all $z, z' \in \bigcup_{i=1}^K \{d_i^k, x_i^k\}$

[6] An alternative model with variable disagreement would be one that seeks to maximize the sum of utilities $u_i(x_i - d_i)$, but this problem is the same as the one studied in Section 4.

$$\begin{cases} U_i(z) - U_i(z') > 0 & \textit{if } z < z' \\ M_i(z')(z - z') \geq U_i(z) - U_i(z') \end{cases}$$

Proposition 6 is straightforward. We simply say that we need numbers $U_i(z)$ to signify levels of utility, and $M_i(z)$ for supergradients, or marginal utilities. The first system of equalities ensure that the first-order conditions for the maximization of the Nash product holds. The second set of inequalities make sure that utility is monotonic and that marginal utilities are a supergradient of the utility.

Proof. We first show that if we are given increasing and concave utility functions u_i, then $x_1^k, \ldots, x_n^k$ is a solution to $\max_{\sum_{i\in N} x_i = M} \prod_{i\in N}[u_i(x_i) - u_i(d_i^k)]$ if and only if for each i, there is a supergradient μ_i of u_i at x_i^k for which

$$\frac{\mu_i}{U_i(x_i^k) - U_i(d_i^k)} = \frac{\mu_j}{U_j(x_j^k) - U_j(d_j^k)}.$$

To this end, define $\mathcal{U} = \{\big(u_1(x_1) - u_1(d_1^k), \ldots, u_n(x_n) - u_n(d_n^k)\big) : \sum_{i\in N} x_i = M, x_i \geq d_i^k\}$. consider maximizing the function $f(y) = \prod_{i\in N} y_i$ subject to $y \in \mathcal{U}$. A point $u \in \mathcal{U}$ maximizes f if and only if $\left(\prod_{i\neq 1} u_i, \ldots, \prod_{i\neq n} u_i\right)$ supports $\mathcal{U}$ at u (by definition). Because f is strictly convex, and since $\mathcal{U}$ is convex and compact, there is a unique such maximizer u^*. It is clear that $u_i^* > 0$ for all $i \in N$.

This states that there is a unique solution $x_1^k, \ldots, x_n^k$ to the Nash problem for which $u_i(x_i^k) - u_i(d_i^k) = u_i^*$. We define $\lambda_j = \prod_{i\neq j}[u_i(x_i^k) - u_i(d_i^k)]$. We know that $\sum_{i\in N} \lambda_i u_i(x_i)$ is maximized at $x_1^k, \ldots, x_n^k$ across all x_i for which $\sum_{i\in N} x_i = M$. Our next step is to show that this can happen if and only if the vector $(1/\lambda_1, \ldots, 1/\lambda_n)$ is proportional to a vector of supergradients.

Since the constraints $x_i \geq d_i^k$ are not binding, we can set up the Lagrangean for the problem, say $L(x, \mu) = \sum_{i\in N} \lambda_i u_i(x_i) + \mu(M - \sum_{i\in N} x_i)$, and note that it is equal to $L(x, \mu) = \sum_{i\in N}[\lambda_i u_i(x_i) - \mu x_i] + \mu M$. We know the constraint $\sum_{i\in N} x_i = M$ is binding so that the solution to the maxmin problem features $\mu^* > 0$. For μ^*, we know that

$\max_x L(x, \mu^*)$ is equal to the maximum Nash product subject to the constraint, and has the same solution. This is equivalent to saying that $(\lambda_i/\mu^*)u_i(x_i^k) - x_i^k \geq (\lambda_i/\mu^*)u_i(x_i) - x_i$ for all x_i, or, rewriting:

$$u_i(x_i) + (\mu^*/\lambda_i)(x_i^k - x_i) \leq u_i(x_i^k).$$

This is equivalent to saying that μ^*/λ_i is a supergradient, or that the vector $(1/\lambda_1, \ldots, 1/\lambda_n)$ is proportional to a supergradient.

Another way of saying that $(1/\lambda_1, \ldots, 1/\lambda_n)$ is proportional to a vector of supergradients is saying that for all $i \in N$, there is a supergradient $M_i(x_i^k)$ of u_i at x_i^k for which for all i, j, $\frac{\lambda_i}{\lambda_j} = \frac{M_j(x_j^k)}{M_i(x_i^k)}$. Writing out the explicit form of λ and eliminating terms, this is equivalent to saying that $\frac{M_i(x_i^k)}{u_i(x_i^k) - u_i(d_i^k)} = \frac{M_j(x_j^k)}{u_j(x_j^k) - u_j(d_j^k)}$, which is precisely the condition in the theorem. The other conditions simply say that M_i is a supergradient, and that u_i is strictly increasing.

Conversely, the details of how to construct a utility function from these numbers essentially follow from Afriat, defining $u_i(x) = \inf_{z \in \bigcup_{k=1}^K \{x_i^k, d_i^k\}} U_i(z) + M_i(z)(x - z)$, where the infimum is taken over all data points. It is then simple to verify by construction that for all $z \in \bigcup_{k=1}^K \{x_i^k, d_i^k\}$, $M_i(z)$ is a supergradient of u_i at z. From this, the fact that the equality in the statement of the theorem is solved implies that the Nash product is maximized for this collection of utility functions (by the previous argument). □

For Proposition 6 to be useful, it must be accompanied by a procedure that one can perform on a data set and decide if the data are Nash rationalizable. Such a procedure is discussed in Chambers and Echenique (2011).

5.3. **The egalitarian model.** Finally, we turn to the egalitarian, or max-min, model. A data set D is egalitarian rationalizable if there are continuous and strictly increasing utility functions $u_i : \mathbf{R}_+ \to \mathbf{R}$ such that, for all k, i and j,

$$\left[u_i(x_i^k) - u_i(d_i^k)\right] = \left[u_j(x_j^k) - u_j(d_j^k)\right].$$

We shall first discuss a strong version of the theory, called the *equal gains* theory, where we require that all agents share the same utility function u. Equal gains is the empirically most interesting version of the theory, because it applies to a test of fairness for the tax code.

We have emphasized that observations (x^k, d^k) should be interpreted as bargaining outcomes, where d^k is a disagreement point, but there are other interpretations. We can, instead think of x^k_i and agent i's pre-tax income and d^k_i his income after taxes. Then, to require that there be some monotonic function u for which,

$$\left[u(x^k_i) - u(d^k_i)\right] = \left[u(x^k_i) - u(d^k_i)\right],$$

says that the tax-code is compatible with all agents sharing equally in the loss of utility derived from taxation. Young (1990) studies the equal gains (or equal loss) model, under the taxation interpretation. Young's empirical results are of a parametric nature. In contrast, we shall present a non-parametric test for the compliance of the tax code with the principle of equal gains.

Our first result assumes only that u is monotonic. Our second result requires that u is also concave (as in the results of Section 4). Below we also discuss the (easy) extension to when utility is allowed to differ across agents.

To begin to understand the empirical content of the equal gains model, let us suppose we have two agents, so that $N = \{1, 2\}$, and that we observe the two data points:

$$d^1 = (0, 7), x^1 = (5, 8) \text{ and } d^2 = (1, 3), x^2 = (2, 8).$$

We claim that this data set cannot be rationalized. To see why, suppose that u were a utility function rationalizing these data using the equal gains model. Then we would have that $u(5) - u(0) = u(8) - u(7)$, and that $u(2) - u(1) = u(8) - u(3)$. Therefore, by adding up the equalities $u(5) - u(0) + (u(7) - u(8)) = 0$ and $u(8) - u(3) + (u(1) - u(2)) = 0$ we must have

$$[u(7) - u(8)] + [u(8) - u(3)] + [u(5) - u(0)] + [u(1) - u(2)] = 0. \tag{1}$$

But we can regroup terms in this expression, obtaining the following:

$$(2)\quad [-u(2)+u(7)]+[-u(8)+u(8)]+[-u(3)+u(5)]+[-u(0)+u(1)] = 0.$$

The contradiction arises because the monotonicity of u implies that each term in brackets in equation (2) is nonnegative, and at least one of them is strictly positive (in fact, each of the terms $[-u(2)+u(7)]$, $[-u(3)+u(5)]$, and $[-u(0)+u(1)]$ are strictly positive). Therefore, the terms cannot add up to zero.

To develop a feeling for what kinds of data are rationalizable, consider a somewhat more involved example. Consider the data:

$$d^1 = (1,8), x^1 = (3,9), d^2 = (2,8), x^2 = (5,9),$$
$$d^3 = (2,9), x^3 = (4,10), d^4 = (0,9), x^4 = (4,10).$$

A rationalizing utility u must the satisfy $u(3) - u(1) = u(9) - u(8)$, $u(5) - u(2) = u(9) - u(8)$, $u(4) - u(2) = u(10) - u(9)$, and $u(4) - u(0) = u(10) - u(9)$. By adding and subtracting, we obtain:

$$([u(1) - u(3)] + [u(5) - u(2)] + [u(2) - u(4)] + [u(4) - u(0)])+$$
$$([u(8) - u(9)] + [u(9) - u(10)] + [u(10) - u(9)] + [u(9) - u(8)]) = 0.$$

But note again, by regrouping, we obtain:

$$([-u(0) + u(1)] + [-u(3) + u(5)] + [-u(2) + u(2)] + [-u(4) + u(4)])+$$
$$([-u(8) + u(8)] + [-u(9) + u(9)] + [-u(10) + u(10)] + [-u(9) + u(9)] = 0.$$

And again, using the monotonicity of u, each of the terms inside of the brackets is nonnegative, and some are strictly positive. This results in a contradiction.

In each of the two examples, what we have done is the following. We have taken data points that, if rationalizable, should force a certain expression to add to zero. By regrouping the terms, the monotonicity of u forces a contradiction; the expression could not possibly add to zero if utility is monotonic. It turns out that the inability to regroup data in this sense is necessary and sufficient for the data to be rationalizable. The inability to regroup data in the appropriate way will be a condition

$$\begin{tikzcd}
 & (8,3) \arrow[dr, "3\leq 5"] & \\
(7,8) \arrow[ur, "8\leq 8"] & & (5,0) \arrow[dl, "1\geq 0"] \\
 & (1,2) \arrow[ul, "7\geq 2"] &
\end{tikzcd}$$

FIGURE 1. A cycle.

(or a nonparametric test) that is equivalent to rationalization by the equal gains model.

In order to formalize our condition, we have to be more specific in what we mean by "regrouping data." It is easiest to think of this in graph theoretic terms. In equation (1), we can think of edges pointing from 7 to 8, from 8 to 3, from 5 to 0, and from 1 to 2. Note that these edges come in "pairs," namely, the edge pointing up from 7 to 8 comes from the data point $(d^1, x^1) = ((0,7),(5,8))$, and is naturally paired with the edge pointing down from 5 to 0. Likewise, the edge pointing up from 1 to 2 is naturally paired with the edge pointing down from 8 to 3.

The interesting point is that when we put these edges together in the appropriate sequence, they form a cycle: see Figure 1. Consider the "edges" $(7,8),(8,3),(5,0),(1,2)$. The endpoints of adjacent edges here are ordered, where we treat $(1,2)$ and $(7,8)$ to be adjacent. That is, the second number (number 8) of the node $(7,8)$ is less than or equal (in fact, equal) to the first number in $(8,3)$. And so forth, for each pair of adjacent edges. In fact, the second number (3) of the node $(8,3)$ is strictly less than the first number (5) in $(5,0)$, as $3 < 5$. And, returning to equation (2), we see that when we regrouped the data, the term $-u(3) + u(5)$ appeared.

We proceed to define our condition formally. First, we define a *cycle* to be a finite sequence of ordered pairs of real numbers, $\{(z_l^1, z_l^2)\}_{l=1}^L$, for which for all $l = 1, \dots, L-1$, $z_l^2 \leq z_{l+1}^1$ and $z_L^2 \leq z_1^1$. A *strict cycle* is a cycle $\{(z_l^1, z_l^2)\}_{l=1}^L$, for which for some l, $z_l^2 < z_{l+1}^1$ or $z_L^2 < z_1^1$. A finite sequence $\{(z_l^1, z_l^2)\}_{l=1}^L$ *defines a (strict) cycle* if there exists a bijection $\sigma : L \to L$ for which $\{(z_{\sigma(l)}^1, z_{\sigma(l)}^2)\}_{l=1}^L$ is a (strict) cycle. Then the ordered pairs $\{(7,8),(8,3),(5,0),(1,2)\}$ from our first example form a strict cycle.

We might conjecture that for data not to be rationalizable, we should be able to pair "up" edges with "down" edges in a way that forms a strict cycle. But this is not quite enough. If we look at the regrouping in the second example, we again paired up edges with down edges. But we did not end up with a single cycle. In fact, we ended up with *two* cycles, only one of which was strict. Namely, the edges

$$\{(1,3),(5,2),(2,4),(4,0),(8,9),(9,10),(10,9),(9,8)\}$$

do not themselves form a cycle, but the two sets of edges $\{(1,3),(5,2),(2,4),(4,0)\}$, $\{(8,9),(9,10),(10,9),(9,8)\}$ each form a cycle. Only the first cycle here is strict, but that is all we need.

In general, we can see there is no reason that a sequence of paired edges need correspond to one, two, or even k cycles. All that we need to obtain a contradiction is that data can be grouped into paired edges which can be partitioned into cycles, at least one of which is strict. These observations motivate the following definitions.

Let L be a natural number, and let $\{(a_l, b_l)\}_{l=1}^L$ and $\{(a'_l, b'_l)\}_{l=1}^L$ be two sequences of L ordered pairs. Say that $\{(a_l, b_l)\}_{l=1}^L$ and $\{(a'_l, b'_l)\}_{l=1}^L$ can be *partitioned into cycles* if there exists a natural number T, and for each $t \leq T$, a collection of finite sequences $\{(z_{tl}^1, z_{tl}^2)\}_{l=1}^{L_t}$ which define cycles (at least one cycle of which is strict), for which there exists a bijection

$$f : \{(t,l) : t \leq T, l \leq L_t\} \to \{(l,i) : l \leq L, i = \{1,2\}\}$$

such that

$$(z_{tl}^1, z_{tl}^2) = \begin{cases} (a_{f_1(t,l)}, b_{f_1(t,l)}) \text{ if } f_2(t,l) = 1 \\ (a'_{f_1(t,l)}, b'_{f_1(t,l)}) \text{ if } f_2(t,l) = 2 \end{cases}$$

The inability to partition paired data points into cycles is exactly the necessary and sufficient condition needed to guarantee that data are rationalizable.

Proposition 7. *The data set* $D = \{(d^k, x^k) : k = 1, \ldots, K\}$ *is rationalizable if and only if there are no sequences* $(d^l, x^l)_{l=1}^L$ *in* D, *and agents* $i_l \neq j_l$ *for all* l, *such that* $([d^l_{i_l}, x^l_{i_l}])_{l=1}^L$ *and* $([x^l_{j_l}, d^l_{j_l}])_{l=1}^L$ *can be partitioned into cycles, at least one of which is strict.*

Two points are worth mentioning. The definition of cycle does not preclude repetition of elements; nor does the notion of "sequence of data points" referred to in the statement of the Proposition. As a result, it may not be obvious how to operationalize the test.

An alternative, computationally viable, test to the condition in Proposition 7 is the following. There is a rationalizing utility if and only if the linear program:

$$\max_{(e,u)\in\mathbf{R}^{|X|+1}} e$$
$$\text{s.t.} \begin{cases} (\forall i,j,k) & ((1_{x^k_i} - 1_{d^k_i}) + (1_{d^k_j} - 1_{x^k_j}))\cdot u \geq 0 \\ (\forall z,z' \in X) & z < z' \Rightarrow (1_{z'} - 1_z)\cdot u > 0. \end{cases}$$

Where $X \subseteq \mathbf{R}^n$ is a finite set such that $d^k, x^k \in X$ for all k.[7]

The relation with the linear program above is obvious from the proof. In fact, the proof of proposition follows from using the following version of a result in linear programming (Farka's Lemma).

Lemma 8. *(Integer-Real Farkas) Let* $\{A_i\}_{i=1}^K$ *be a finite collection of vectors in* $\mathbf{Q}^n$. *Then one and only one of the following statements is true:*

i) *There exists* $y \in \mathbf{R}^n$ *such that for all* $i = 1, ..., L$, $A_i \cdot y \geq 0$ *and for all* $i = L+1, ..., K$, $A_i \cdot y > 0$.
ii) *There exists* $z \in \mathbf{Z}^K_+$ *such that* $\sum_{i=1}^K z_i A_i = 0$, *where* $\sum_{i=L+1}^K z_i > 0$.

Proof. B i) and ii) cannot simultaneously hold. To see why, suppose that there exist y and z as stated. Then $A_i \cdot y \geq 0$ for all $i = 1, \ldots, L$

[7] We thank an anonymous referee for suggesting this formulation of the argument.

and $A_i \cdot y > 0$ for all $i = L+1, \ldots, K$. Consider $\sum_{i=1}^{K} z_i A_i \cdot y$. Since $\sum_{i=1}^{K} z_i A_i = 0$, we know that $\sum_{i=1}^{K} z_i A_i \cdot y = 0$. Furthermore, since there is some $j \in \{L+1, \ldots, K\}$ for which $z_j A_j \cdot y > 0$, and for all i, $z_i A_i \cdot y \geq 0$, we conclude that $\sum_{i=1}^{K} z_i A_i \cdot y > 0$, a contradiction.

We now establish that if ii) does not hold, i) holds. By Theorem 3.2 of Fishburn (1973), if ii) does not hold, there exists $y \in \mathbf{Q}^n$ such that for all $i = 1, ..., L$, $A_i \cdot y \geq 0$ and for all $i = L+1, ..., K$, $A_i \cdot q > 0$. Since $\mathbf{Q}^n \subseteq \mathbf{R}^n$, $y \in \mathbf{R}^n$. $\square$

5.4. **Proof of Proposition 7.** Let $X \subseteq \mathbf{R}^n$ be a finite set such that $d^k, x^k \in X$ for all k.

The notation 1_x refers to a vector of zeroes and ones, with a one in the x coordinate and zero elsewhere (an indicator function).

There is a rationalizing u if and only if there is a solution to the system of linear inequalities:

$$(\forall i, j, k) \qquad ((1_{x_i^k} - 1_{d_i^k}) + (1_{d_j^k} - 1_{x_j^k})) \cdot u \geq 0 \tag{3}$$

$$(\forall z, z' \in X) \qquad z < z' \Rightarrow (1_{z'} - 1_z) \cdot u > 0. \tag{4}$$

Thus, there is an inequality (3) for each i, j and k, and an inequality (4) for each $z', z \in X$ with $z < z'$.

Once a solution to the linear inequalities has been obtained, the function u can be completed by linear interpolation.

By Lemma 8, there is no solution to system (3)-(4) if and only if there are vectors $\lambda \in \mathbf{Z}_+^{KN^2}$ and $\theta \in \mathbf{Z}_+^{|X|^2}$ with

$$\sum_{k,i,j} \lambda_{k,i,j}((1_{x_i^k} - 1_{d_i^k}) + (1_{d_j^k} - 1_{x_j^k})) + \sum_{(z,z'):z'>z} \theta_{z,z'}(1_{z'} - 1_z) = 0$$

and $\sum_{(z,z'):z'>z} \theta_{z,z'} > 0$.

Without loss of generality, we can assume that $d_i^k \neq d_j^k$ and $x_i^k \neq x_j^k$ for all k and all $i \neq j$. To see this note that, if $d_i^k = d_j^k$ then $x_i^k < x_j^k$ implies that there is no rationalizing monotonic u; but then the intervals $[d_i^k, x_i^k]$ and $[d_j^k, x_j^k]$ define a strict cycle: $\{(d_i^k, x_i^k)\}$ $\{(x_j^k, d_j^k)\}$. Similarly if $x_j^k < x_i^k$. On the other hand, $x_i^k = x_j^k$ implies that the inequalities corresponding to k, i, j in (3) are always satisfied. So these inequalities are irrelevant to the existence of a rationalizing u. The argument is analogous when $d_i^k \neq d_j^k$ and $x_i^k = x_j^k$.

Let the vectors $(\lambda_{k,i,j})$ and $(\theta_{z,z'})$ be as above. Consider the following collections of vectors in $\{-1,0,1\}^X$: Let A_D be the collection of vectors with $\lambda_{k,i,j}$ copies of $(1_{d_j^k} - 1_{x_j^k})$; let A_U be the collection with $\lambda_{k,i,j}$ copies of $(1_{x_i^k} - 1_{d_i^k})$. Let $f : A_D \to A_U$ be the bijection which associates each $(1_{d_j^k} - 1_{x_j^k})$ with a different copy of $(1_{x_i^k} - 1_{d_i^k})$. Such a bijection exists given the way that A_D and A_U were constructed.

Let A_M be the collection with $\theta_{z,z'}$ copies of $1_{z'} - 1_z$ for each $z, z' \in X$ with $z' > z$. By definition of λ and θ, we know that the sum of the elements of A_D, A_U, and A_M equals the null vector. We also have $A_M \neq \emptyset$.

Let $G = (X, E)$ be the graph obtained by letting there be an edge pointing from x to x' if and only if there is a vector $1_{x'} - 1_x$ in one of the collections A_D, A_U or A_M. By the Poincaré-Veblen-Alexander Theorem (see Berge (2001), p. 148, Theorem 5), since the sum of the elements of the vectors in A_D, A_U, and A_M equals the null vector, G can be partitioned into circuits $C_1, \dots C_T$. Note that, if $e = (v, v') \in A_U \cup A_M$ then $v \leq v'$. If $e = (v, v') \in A_D$, then $v \geq v'$.

Consider the edges in circuit C_t: Let $\left[d_{i_l}^l, x_{i_l}^l\right]$, $l = 1, \dots L_t^U$ be the set of intervals defined by edges $(d_{i_l}^l, x_{i_l}^l) \in A_U$ and $\left[d_{j_l}^l, x_{j_l}^l\right]$ $l = 1, \dots L_t^D$ be the set of intervals defined by edges $(x_{j_l}^l, d_{j_l}^l) \in A_D$. For any edge $e = (v, v') \in A_U \cup A_D$ in C_t, let (v'', v''') be the first edge in C_t after e that is in $A_U \cup A_D$. Then either $v' = v''$ or there are edges in A_M between e and (v'', v''') in C_t; so $v' \leq v''$. Hence, for any $e = (v, v') \in A_U \cup A_D$ in C_t, the successor edge $(v'', v''') \in A_U \cup A_D$ satisfies that $v' \leq v''$. Hence the intervals $(d_{i_l}^l, x_{i_l}^l)$ $l = 1, \dots L_t^U$ and $(x_{j_l}^l, d_{j_l}^l)$ $l = 1, \dots L_t^D$ define a cycle.

In addition, since $A_M \neq \emptyset$, at least one of the sets of intervals defined by a circuit C_t defines a strict cycle.

Finally, since there is a bijection between the edges in A_U and in A_D, we have $\sum_t L_t^U = \sum_t L_t^D = L$. So if we let $(\left[d_{i_l}^l, x_{i_l}^l\right])_{l=1}^L$ collect the sequences $\left[d_{i_l}^l, x_{i_l}^l\right]$, $l = 1, \dots L_t^U$, and $(\left[x_{j_l}^l, d_{j_l}^l\right])_{l=1}^L$ collect the sequences

$\left[d^l_{j_l}, x^l_{j_l}\right]$ $l = 1, \ldots L^D_t$, then we have a sequence of intervals in the condition in the statement of the proposition.

5.5. **Strengthening.** While the condition above derives necessary and sufficient conditions for rationalizability by *some* increasing utility function, most natural interpersonally comparable utility functions are assumed to be concave, to account for diminishing marginal utility. To this end, we ask what are the additional testable implications of requiring the rationalizing u to be concave?

The answer is not so difficult and is closely related to work of Afriat (1967), Richter and Wong (2004), and Kalandrakis (2010). To simplify matters, we suppose in this section that all observed data points are rational. Thus, say that a data set D is *rational-valued* if all d^k_i and all x^k_i are elements of $\mathbf{Q}$.

We will say that a data set D is *concave rationalizable* if there is a strictly increasing, continuous, concave $u : \mathbf{R}_+ \to \mathbf{R}$ for which for all k, i and j,

$$[u(x^k_i) - u(d^k_i)] = [u(x^k_j) - u(d^k_j)].$$

Our first task is to describe an example whereby concave rationalizability fails. It is easy to construct such an example: consider the two agent case, and the one point data set

$$d^1 = (0, 2), x^1 = (2, 3).$$

How can we see that there is no concave u which concave rationalizes this data set? Clearly, if a concave u exists, it must be that $u(2) \geq (2/3)u(3) + (1/3)u(0)$, or $3u(2) \geq 2u(3) + u(0)$. This expression is obviously equivalent to $2[u(2) - u(3)] + [u(2) - u(0)] \geq 0$. Finally, we know that $[u(3)-u(2)]+[u(0)-u(2)] = 0$, so by adding the two terms, we obtain $[u(2) - u(3)] \geq 0$, which we know to be a contradiction to monotonicity.

If we think in the context of the previous section, what we are doing is adding new types of "edges" to our graph. In the previous section, we could add an edge from d^k_i to x^k_i so long as we added a corresponding edge from some x^k_j to d^k_j. Now, we are also allowed to add certain "collections" of edges; namely, any collection $\{(a_1, b), ..., (a_n, b)\}$ for which $nb = \sum_{l=1}^{n} a_l$; that is, b is a rational convex combination of the a_l terms. This is precisely what we did in the previous example. We have the collection of edges $(3, 2), (3, 2), (0, 2)$, which comes from the fact that $2 = (2/3)3 + (1/3)0$. To this, we add the edges $(2, 3), (2, 0)$, which comes from the fact that we have equal gains. Combining these together results in the cycles

$$(3, 2), (2, 3),$$

$$(0, 2), (2, 0),$$

and

$$(3, 2).$$

Note that the singleton element $(3, 2)$ is by itself a strict cycle.

To this end, we discuss the following generalization of the concept from the previous section. We will say a collection of ordered pairs $\{(a_1, b), ..., (a_m, b)\}$ is a *convex collection* if $\sum_{o=1}^{m} a_o = mb$. That is, b is a convex combination of the a_o terms. Note that we do not preclude the possibility of several a_o terms being equal.

Let P be a natural number. For each $p \leq P$, let L_p be a natural number, and let $\{(a_l^p, b_l^p)\}_{l=1}^{L_p}$ be a sequence of ordered pairs. Say the sequences that $\{(a_l^p, b_l^p)\}_{l=1}^{L_p}$ can be *partitioned into cycles* if there exists a natural number T and for each $t \leq T$, a collection of finite sequences $\{(z_{tl}^1, z_{tl}^2)\}_{l=1}^{L_t}$ which define cycles (at least one cycle of which is strict), for which there exists a bijection $f : \{(t, l) : t \leq T, l \leq L_t\} \to \{(l, i) : l \leq L_i, i \in \{1, \ldots, P\}\}$ for which $(z_{tl}^1, z_{tl}^2) = (a_{f_1(t,l)}^p, b_{f_1(t,l)}^p)$ if $f_2(t, l) = p$.

Proposition 9. *The data $D = \big\{(d^k, x^k) : k = 1, \ldots, K\big\}$ are rationalizable if and only if there are no sequences of data points $(d^l, x^l)_{l=1}^{L}$ in D, and agents $i_l \neq j_l$ for all l, and Q convex collections $\{(a_o^q, b^q)\}_{o=1}^{m_q}$, each of whose elements is either a d_i^k or x_i^k, such that $([d_{i_l}^l, x_{i_l}^l])_{l=1}^{L}$, $([x_{j_l}^l, d_{j_l}^l])_{l=1}^{L}$, and $\{(a_o^q, b^q)\}_{o=1}^{m_q}$ can be partitioned into cycles, at least one of which is strict.*

5.6. **Proof of Proposition 9.** Let $X \subseteq \mathbf{R}^n$ be a finite set such that $d^k, x^k \in X$ for all k. We introduce two copies of X, one whose indicator functions are written in the standard way (1_x), the other of which is written $1'_x$.

There is a rationalizing u if and only if there is a solution to the system of linear inequalities in the X dimensional variables u and α,

$$((1_{x_i^k} - 1_{d_i^k}) + (1_{d_j^k} - 1_{x_j^k})) \cdot (u, \alpha) \geq 0 \tag{5}$$

$$(1_{z'} - 1_z) \cdot (u, \alpha) > 0 \tag{6}$$

$$(1_{z'} - 1_z) + (z - z')1'_{z'} \cdot (u, \alpha) \geq 0 \tag{7}$$

where equation (6) runs across $z' > z$.

To see why, suppose there is a rationalizing u. Without loss, we may suppose that u is piecewise linear (if u concave rationalizes the data, then so does the piecewise linear function which takes the same values as u for every x_i^k and d_i^k). Then u has a supergradient α at every x_i^k and d_i^k. This is the content of equation (7). The other two inequalities are obviously satisfied.

Conversely, suppose the three equations are satisfied. First, we claim that without loss, each α term is greater than zero. For example, if we consider α_x where $x < y$ for some $y \in X$, then by equation (7), it follows that $\alpha_x(y - x) + u_x - u_y > 0$, or $\alpha_x > \frac{u_y - u_x}{y - x}$ (since $u_y > u_x$ by equation (6)). Further, if there is no $y \in X$ for which $y > x$, then we can always redefine $\alpha_x = \min_{y \in X, x \neq y} \frac{u_y - u_x}{y - x} > 0$, which results in another system of consistent weights.

Then it is a standard trick, due to Afriat (1967), to define

$$u(y) = \min_{x \in X} u_x + \alpha_x(y - x),$$

and note that this function is concave, strictly monotonic, and rationalizes the data.

Now, we can sketch the argument as to why the satisfaction of the system of inequalities is equivalent to the absence of cycles as stated in Proposition 9. It is again an application of Lemma 8. Equations (5) through (7) have no solution if and only if there are there are vectors $\lambda \in \mathbf{Z}_+^{KN^2}$, $\theta \in \mathbf{Z}_+^{|X|^2}$, and $\eta \in \mathbf{Z}_+^{|X|^2}$, with

$$\sum_{k,i,j} \lambda_{k,i,j}((1_{x_i^k} - 1_{d_i^k}) + (1_{d_j^k} - 1_{x_j^k}))$$
$$+ \sum_{(z,z'):z'>z} \theta_{z,z'}(1_{z'} - 1_z) + \sum_{(z,z')} \eta_{z,z'}(1_{z'} - 1_z) + (z - z')1'_{z'} = 0$$

and $\sum_{(z,z'):z'>z} \theta_{z,z'} > 0$. Importantly, from this equation we can infer that the two equations:

$$\sum_{k,i,j} \lambda_{k,i,j}((1_{x_i^k} - 1_{d_i^k}) + (1_{d_j^k} - 1_{x_j^k})) + \sum_{(z,z'):z'>z} \theta_{z,z'}(1_{z'} - 1_z) + \sum_{(z,z')} \eta_{z,z'}(1_{z'} - 1_z) = 0$$

and

$$\sum_{(z,z')} \eta_{z,z'}(z - z')1'_{z'} = 0 \tag{8}$$

are jointly satisfied.

The proof now proceeds in the same way as the proof of Proposition 7. There is only one change. Now, we also consider a collection of X dimensional vectors $\{-1, 0, 1\}$, which we call A_C, which consists of $\eta_{z,z'}$ copies of each $1_z - 1_{z'}$. The graph $G = (X, E)$ is now constructed in the same way as in the proof of Proposition 7, letting there be one edge from x to x' for each copy of $1_{x'} - 1_x$ in A_D, A_U, A_M or A_C. Now, consider equation (8). This equation implies in particular that for all z', $\sum_z \eta_{z,z'}(z - z') = 0$, in other words, the collection of edges pointing to z' in A_C form a concave collection.

5.7. **Generalization.** We have here asked for data to be rationalized by a single utility function, common to all $i \in N$. If, instead, we ask that for each i, there exists $u_i : \mathbf{R} \to \mathbf{R}$ for which for all $i, j \in N$

$$u_i(x_i^k) - u_i(d_i^k) = u_j(x_j^k) - u_j(d_j^k),$$

we obviously get a weaker condition. The weakening required here is simply that when partitioning data into cycles, each cycle can only contain edges corresponding to a *single* agent. The proof is similar to the proceeding and is hence omitted.

5.8. **An application to spatial competition.** The result in Proposition 7 has a simple application to the testable implications of Hotelling's model of spatial competition (Hotelling (1929)). Hotelling's model concerns the location of two vendors on a unidimensional space, and a distribution of consumers. Consumers always buy from the vendor closest to them (in the case of equidistant vendors, half of the consumers go to one vendor, and the other half to the other), and each vendor's profit consists of how many consumers buy from him. Hotelling's model is not about bargaining, but it seems potentially useful to point out the application of our results.

In our version of Hotelling's model, we observe a finite collection of intervals $([a^k, b^k]) \subseteq [0,1]$, and for each observed interval, a location $m^k \in (a^k, b^k)$. We want to know, when does there exist a full-support distribution μ of agents on $[0,1]$ such that for each k, m^k is the median of μ conditional on $[a^k, b^k]$? This provides us with the testable implications of the Hotelling model when the distribution of agents is unobserved, but when the boundaries of spatial competition can vary.

The relation to section 5 is as follows. A distribution μ satisfying the properties exists if and only if there is a strictly increasing $F : [0,1] \to \mathbb{R}$ (a cdf) for which for all k, $F(b^k) - F(m^k) = F(m^k) - F(a^k)$. Now, imagine that in the previous section we had only two agents ($|N| = 2$), and $d^k = (m^k, a^k)$, $x^k = (b^k, m^k)$.

This leads us directly to the following corollary:

Corollary 10. *A finite list of intervals* $[a^k, b^k]$ *and locations* m^k *is consistent with the Hotelling model if and only if there are no sequences of data points* $[a^l, b^l]_{l=1}^{L}$, $[a^l, b^l]_{l=L+1}^{L'}$ *for which* $\{(a^l, m^l)\}_{l=1}^{L}$, $\{(b^l, m^l)\}_{l=1}^{L}$, $\{(m^l, a^l)\}_{l=L+1}^{L'}$, $\{(m^l, b^l)\}_{l=L+1}^{L'}$ *can be partitioned into cycles.*

6. Conclusion

We consider finite sets of observations of bargaining outcomes, and develop the testable implications of some of the best-known models in bargaining theory.

We consider two basic frameworks. Our results are sharpest for the case where we assume that disagreement points are fixed across observations. We show that the utilitarian, Nash bargaining, and egalitarian max-min models are all observationally equivalent. Further, we show that a simple test for these models consists in checking that the observed allocations are comonotonic.

When disagreement points can vary, we present a characterization of the data that are consistent with a form of egalitarianism, namely the model of equal gains/losses. The main application of these results are to tax data, where we can check for consistency of the tax code with the principle of equal loss when the utility function is unknown.

References

Afriat, Sydney N., "The construction of utility functions from expenditure data," *International Economic Review*, 1967, *8* (1), 67–77.

Berge, Claude, *The theory of graphs*, Dover Publications, 2001.

Bowlby, Roger L. and William R. Schriver, "Bluffing and the "split-the-difference" theory of wage bargaining," *Industrial and Labor Relations Review*, 1978, *31* (2), 161–171.

Carvajal, Andres and Natalia González, "On refutability of the Nash bargaining solution," 2011. mimeo, University of Warwick.

Chambers, Christopher P. and Federico Echenique, "Testable Implications of Bargaining Theories," 2011. Caltech SS Working Paper 1348.

Cherchye, Laurens, Thomas Demuynck, and Bram. De Rock, "Nash bargained consumption decisions: A revealed preference analysis," 2011. CES-Discussion paper series DPS11. 07.

Chiappori, Pierre A., Olivier Donni, and Ivana Komunjer, "Learning from a piece of pie," *Review of Economic Studies*, 2012, pp. 162–195.

Clippel, Geoffroy De and Kfir Eliaz, "Reason-Based Choice: A Bargaining Rationale for the Attraction and Compromise Effects," 2011. Forthcoming, Theoretical Economics.

Fishburn, Peter C., *The theory of social choice*, Princeton University Press Princeton, NJ, 1973.

Hamermesh, Daniel S., "Who 'wins' in wage bargaining?," *Industrial and Labor Relations Review*, 1973, pp. 1146–1149.

Hotelling, Harold, "Stability in competition," *The economic journal*, 1929, *39* (153), 41–57.

Kalai, Ehud, "Proportional solutions to bargaining situations: interpersonal utility comparisons," *Econometrica*, 1977, pp. 1623–1630.

Kalandrakis, Tasos, "Rationalizable voting," *Theoretical Economics*, 2010, *5* (1), 93–125.

Kıbrıs, Özgür., "A revealed preference analysis of solutions to simple allocation problems," *Theory and Decision*, 2011, pp. 1–15.

Lensberg, Terje, "Stability and collective rationality," *Econometrica*, 1987, pp. 935–961.

Nash, John F., "The bargaining problem," *Econometrica*, 1950, *18* (2), 155–162.

Rawls, John, *A theory of justice*, Belknap Press, 1971.

Richter, Marcel K. and Kim C. Wong, "Concave utility on finite sets," *Journal of economic theory*, 2004, *115* (2), 341–357.

Roth, Alvin E., "Introduction to Experimental Economics," in John H. Kagel and Alvin E. Roth, eds., *The Handbook of Experimental Economics*, Princeton University Press, 1995, chapter 1.

Sprumont, Yves, "Monotonicity and Solidarity Axioms in Economics and Game Theory," in Prasanta Pattanaik, ed., *Rational Choice and Social Welfare*, Springer, 2008, pp. 71–94.

Svejnar, Jan, "On the empirical testing of the Nash-Zeuthen bargaining solution," *Industrial and Labor Relations Review*, 1980, *33* (4), 536–542.

Thomson, William, "Nash's bargaining solution and utilitarian choice rules," *Econometrica*, 1981, pp. 535–538.

———, "Bargaining theory: the axiomatic approach," 2010. book manuscript.

——— **and Terje Lensberg**, *Axiomatic theory of bargaining with a variable number of agents*, Cambridge Univ Press, 1989.

Young, H. Peyton, "Distributive justice in taxation," *Journal of Economic Theory*, 1988, *44* (2), 321–335.

———, "Progressive taxation and equal sacrifice," *The American Economic Review*, 1990, *80* (1), 253–266.