KATHOLIEKE UNIVERSITEIT LEUVEN

Testable implications of general equilibrium models: an integer programming approach

by

Laurens CHERCHYE
Thomas DEMUYNCK
Bram DE ROCK

Public Economics

DISCUSSION PAPER

Testable implications of general equilibrium models: An integer programming approach[*]

Laurens Cherchye,[†] Thomas Demuynck[‡] and Bram De Rock[§]

July 6, 2009

Abstract

Focusing on the testable implications on the equilibrium manifold, we show that the rationalizability problem is NP-complete. Subsequently, we present an integer programming (IP) approach to characterizing general equilibrium models. This approach avoids the use of the Tarski-Seidenberg algorithm for quantifier elimination that is commonly used in the literature. The IP approach naturally applies to settings with any number of observations, which is attractive for empirical applications. In addition, it can easily be adjusted to analyze the testable implications of alternative general equilibrium models (that include, e.g., public goods, externalities and/or production). Further, we show that the IP framework can easily address recoverability questions (pertaining to the structural model that underlies the observed equilibrium behavior), and account for empirical issues when bringing the IP methodology to the data (such as goodness-of-fit and power). Finally, we show how to develop easy-to-implement heuristics that give a quick (but possibly inconclusive) answer to whether or not the data satisfy the general equilibrium models.

JEL Classification: C60, D10, D51

Keywords: General equilibrium, equilibrium manifold, exchange economies, production economies, NP-completeness, nonparametric restrictions, GARP, integer programming.

1   Motivation

We introduce an integer programming (IP) approach to nonparametrically characterizing general equilibrium models. Attractively, this approach naturally deals with any number of observations and/or agents. We show that we can easily integrate several extensions into our

---

basic model that focus on public goods, externalities and production. In addition, we address recoverability questions and empirical issues such as power and goodness-of-fit of general equilibrium models. This paves the way for nonparametric analysis of general equilibrium behavior in empirical applications.

The Sonnenschein-Mantel-Debreu result can be summarized in the following way: any real valued function of prices that satisfies Walras' law, continuity and homogeneity of degree zero is the excess demand function of some economy with at least as many agents as commodities. This celebrated result led to the rather depressing viewpoint that general equilibrium is unable to generate falsifiable predictions. From a Popperian perspective, this would label general equilibrium theory as unscientific.

Brown and Matzkin (1996), however, showed that if we focus on the equilibrium manifold,[1] and not on the excess demand function, then the pure exchange model has strong nonparametric empirical restrictions. Their main results characterize the finite data sets consisting of equilibrium prices, aggregate endowments and individual incomes, for which there exist continuous, concave and non-satiated utility functions such that the observed prices correspond to some equilibrium price vector for the pure exchange economy associated with the given endowments. More precisely, these utility functions exist if and only if there exist individual consumption bundles such that: (i) individual expenditure equals individual income, (ii) individual consumption bundles sum to aggregate endowments, and (iii) for each individual there exist a solution for the corresponding Afriat inequalities. Moreover, they demonstrated that these restrictions are non-vacuous. If a given data set satisfies the conditions (i), (ii) and (iii) this data set is said to be rationalizable.

Requirements (i) and (ii) are expressed as linear equalities and can therefore easily be verified by linear programming methods. Unfortunately, requirement (iii) has a quadratic form. In order to circumvent this problem, Brown and Matzkin make use of a deep result from semi-algebraic theory: the Tarski-Seidenberg theorem. This theorem states that every first-order formula over the real fields can be reduced to an equivalent quantifier-free formula. Moreover, this reduction can be established in finite time. From this, Brown and Matzkin conclude that it is possible to derive in finite time whether a solution to (i), (ii) and (iii) exists.

Subsequent research has extended the result to models including production (Carvajal, 2005), Pareto optimal provision of public goods (Snyder, 1999), financial markets (Kübler, 2003), random preferences (Carvajal, 2003), Pareto efficient and individual rational allocations (Bachmann, 2006) and models with interdependent preferences (Deb, 2009).[2] The usual approach adopted in these studies is as follows. First, it is demonstrated that there exist concave, continuous and non-satiated utility functions that rationalize the data for the economy under consideration if and only if there exists some set of variables satisfying a certain collection of polynomial inequalities. Second, it is inferred, making use of the Tarski-Seidenberg theorem on quantifier elimination, that the issue of rationalizability can be resolved in finite time. Third, a counterexample is provided, affirming the non-triviality of the collection of derived

---

[1]The equilibrium manifold is the set of prices and individual endowments for which the excess demand function is zero.

[2]We refer to Carvajal et al. (2004) for an extensive overview of the literature.

polynomial inequalities.

Although these results clearly demonstrate the conditions for which a given data set can belong to the equilibrium manifold, most researches remain quite negligent on the issue concerning the practical verification of these conditions. Two exceptions are worth mentioning. In their original paper, Brown and Matzkin (1996) show how to use the Tarski-Seidenberg algorithm in order to derive the testable restrictions for general equilibrium models with 2 agents and 2 observations. Unfortunately, the Tarski-Seidenberg algorithm is, for worst time complexity, at best doubly exponential in the number of quantifiers to be eliminated. Hence, it is computationally very inefficient even for moderate sized problems. A second algorithm was proposed by Brown and Kannan (2008). This algorithm enumerates every possible preference ordering of all individuals over the different observations and verifies for each preference profile —via linear programming techniques— whether these preferences lead to a rationalization. This algorithm is exponential in both the number of observations and the number of individuals. The computational inefficiency of these two algorithms raises the question whether there exist an algorithm that can verify the rationalizability question efficiently, i.e. in a polynomial number of steps. In Section 3, we show that, unless P= NP, the answer is no. In particular, the verification of the empirical restrictions of the Brown and Matzkin characterization is an NP-complete problem.[3] This result implies that one should not try to construct a polynomial time algorithm that verifies the rationalizability question —unless one has taken up the ambitious task of showing that P= NP. In turn, it gives a strong argument in favor of searching for a widely applied and 'efficient' non-polynomial time algorithm, to verify the rationalizability conditions and to open the way for introducing heuristics that can give quick (but possible inconclusive) answers.

By exploiting the equivalence between the existence of Afriat inequalities and the Generalized Axiom of Revealed Preference,[4] we show how to transform condition (iii) of Brown and Matzkin into a set of linear restrictions with mixed integer variables; i.e. we apply an Integer Programming (IP) procedure to characterize testable implications of general equilibrium models. Such an IP approach has proven very useful in the closely related literature of collective consumption models, which studies the behavior of multi-person households; Chiappori et al. (2004) cite multi-person household behavior as a prime instance of 'small group behavior' that can be analyzed by using insights from general equilibrium theory. See Cherchye et al. (2007) and Cherchye et al. (2008) for IP characterizations of collective consumption models.

From a theoretical point of view, our core motivation for adopting an IP approach is that the IP procedure is a widely accepted and a well known tool to handle NP-complete problems. Besides this, we also have a number of other motivations for our IP approach. Most notably, the IP approach allows us to avoid the use of the Tarski-Seidenberg algorithm to eliminate the quantifiers. Although in theory this algorithm can handle data sets with any number of observations, existing applications in the context of general equilibrium models mostly restrict their analysis to data sets with only 2 observations (see our above discussion). Moreover, when using the Tarski-Seidenberg algorithm, the analysis for one general equilibrium model is not

---

[3]We refer to Garey and Johnson (1979) for an introduction into the theory of NP-completeness.

[4]See Section 2 for formal definitions.

straightforwardly extended to another model (that accounts, e.g., for public goods, externalities and/or production). Given this, an important argument pro our IP approach is that it provides a versatile framework for analyzing testable implications of general equilibrium models.[5] For example, in Sections 3 and 4 our characterizations of general equilibrium models *à la* Brown and Matzkin naturally apply to any number of observations. Next, in Sections 5 and 6 we present fairly straightforward extensions of the basic IP characterizations towards other general equilibrium models, and we show how our IP approach can account for empirical issues like power and goodness of fit.

The fact that the IP characterizations developed in the following sections naturally apply to any number of observations, makes them particularly attractive for empirical applications. In this respect, from elementary IP theory, we know that solving integer programs may become computationally demanding for large data sets. However, for such large data sets we illustrate the possible use of heuristics that give a quick (but possibly inconclusive) answer to whether or not the data set satisfies the general equilibrium restrictions.

Finally, our results also provide some useful extensions to the literature concerning the testable implications of household decision making. For example, we propose heuristics that can also be used in the context of collective consumption models. In addition, our IP character-izations of general equilibrium models with production, can serve as a starting point to develop IP characterizations of collective consumption models that account for household production (compare with Chiappori (1997)).

Summarizing, the IP approach that is presented in this paper provides a useful tool for bringing the testable implications of general equilibrium models to the data. In such applica-tions, it can be useful to think of different agents of the general equilibrium model as agent types, representing groups of consumers with the same tastes and incomes; see Brown and Matzkin (1996) for discussion. Next, Chiappori et al. (2004) claim that general equilibrium analysis does not only apply to 'large economies', but also to 'small group behavior' of house-holds, committees, clubs, villages and other local organization. See also Rizvi (2006), and Brown and Kannan (2008) for more discussion on meaningful empirical applications of gen-eral equilibrium models. Finally, Brown and Lee (2008) illustrate that our insights could also be useful to obtain a more accurate measurement of the social cost of monopoly.

The paper unfolds as follows. Section 2 sets the stage by briefly recapturing Afriat (1967)'s theorem, which characterizes rationalizable behavior in terms of Afriat inequalities, and by introducing the Generalized Axiom of Revealed Preference (GARP). Section 3 presents the NP-completeness result, and provides an IP characterization of general equilibrium behavior in a pure exchange economy with assignable incomes; this is the case that was originally con-sidered by Brown and Matzkin (1996). Section 4 looks at the more general case that does not assume fully assignable incomes; the case with partially assignable incomes has not yet been investigated in the literature on general equilibrium models. This section also illustrates the construction of quick heuristics that provide sufficient conditions of rationalizability. Section 5 provides IP characterizations for equilibrium models with public goods and externalities.

---

[5]See Sethuraman et al. (2003) for a similar argument that motivates integer programming characterizations in the context of Arrovian social welfare functions.

4

Section 6 discusses some further extensions of our IP characterizations: (1) incorporate production in the general equilibrium models; (2) address recoverability issues; and (3) account for empirical issues such as goodness-of-fit and power in applications.

2   Preliminaries

We consider a setting with J goods, and a finite data set $S = \{p_t, q_t\}_{t=1,\dots T}$ existing of J-dimensional price vectors $p_t \in \mathbb{R}^J_{++}$ and J-dimensional quantity vectors $q_t \in \mathbb{R}^J_+$. The following definition is a standard concept.

Definition 1. *A data set* $\{p_t, q_t\}_{t=1,\dots,T}$ *is rationalizable by the function* $u : \mathbb{R}^J_+ \to \mathbb{R}$ *if for all* $t = 1, \dots, T$:

$$q_t \in \arg \max_{\{\langle p_t, q \rangle \leq \langle p_t, q_t \rangle\}} u(q).$$

In what follows, we will mainly focus on the nonparametric Generalized Axiom of Revealed Preferences (GARP); Varian (1982) has shown that GARP is a necessary and sufficient condition for the data set $\{p_t, q_t\}_{t=1,\dots T}$ to be rationalizable.

Definition 2. *The data* $\{p_t, q_t\}_{t=1,\dots,T}$ *satisfies* GARP *if and only if we can construct relations* R_o, R *such that*

(i) for all $t, s \in \{1, \dots, T\}$, *if* $\langle p_t, q_t \rangle \geq \langle p_t, q_s \rangle$ *then* $q_t \, R_o \, q_s$;

(ii) for all $t, s, u, \dots, r, v \in \{1, \dots, T\}$, *if* $q_t \, R_o \, q_s$, $q_s \, R_o \, q_u$, $\dots$, *and* $q_r \, R_o \, q_v$ *then* $q_t \, R \, q_v$;

(iii) for all $t, s \in \{1, \dots, T\}$, *if* $q_t \, R \, q_s$, *then* $\langle p_s, q_s \rangle \leq \langle p_s, q_t \rangle$.

Condition (i) states that the quantities q_t are directly revealed preferred over q_s ($q_t \, R_o \, q_s$) if q_t was chosen when q_s was equally attainable ($\langle p_t, q_t \rangle \geq \langle p_t, q_s \rangle$); q_t are strict directly revealed preferred over q_s if the strict inequality holds. Similarly, we will also often state that the individual directly revealed prefers observation t over s. Next, condition (ii) imposes transitivity on the revealed preference relation R. Finally, condition (iii) states that if a consumption bundle q_t is revealed preferred to a consumption bundle q_s, then q_s cannot be more expensive then q_t.

The following well-known theorem by Afriat (1967) states the conditions that a data set has to satisfy in order to be rationalizable.

Theorem 1. *Let* $S = \{p_t, q_t\}_{t=1,\dots,T}$ *be a set of observations. The following statements are equivalent:*

(i) There exist a non-satiated utility function that rationalizes S;

(ii) S satisfies GARP;

(iii) There exist numbers $\{\phi_t\}_{t=1,\dots,T} \geq 0$ and $\{\lambda_t\}_{t=1,\dots,T} > 0$ such that for all $t, s \in \{1,\dots,T\}$:

$$\phi_t \leq \phi_s + \lambda_s \left(\langle p_s, q_t \rangle - \langle p_s, q_s \rangle \right);$$

(iv) There exist a non-satiated, continuous, concave and monotonic utility function that rationalizes S.

Condition (ii) recaptures the result of Varian (1982). Condition (iii) provides an equivalent characterization in terms of the Afriat inequalities, which allow an explicit construction of the utility levels associated with each observation t (i.e., utility level ϕ_t for observed quantities q_t). Finally, condition (iv) states that, if the data set is rationalizable by some utility function u, then it is also rationalizable by a well-behaved (i.e., continuous, monotonic and concave) utility function.

We end this section by providing a fifth equivalent to the above theorem, i.e. an integer programming (IP) characterization of GARP. To do so, we introduce the binary variables $x_{s,t} \in \{0, 1\}$. Consider then the following set of constraints:

Program (CS.I).

$$
\begin{align}
(i) & \quad \langle p_t, q_t \rangle - \langle p_t, q_s \rangle < x_{t,s}\, A_t & (t, s = 1, \dots, T); \\
(ii) & \quad x_{t,s} + x_{s,v} \leq 1 + x_{t,v} & (t, s, v = 1, \dots, T); \\
(iii) & \quad (x_{t,s} - 1)\, A_s \leq \langle p_s, q_t \rangle - \langle p_s, q_s \rangle & (t, s = 1, \dots, T).
\end{align}
$$

Where the constants A_t are given real numbers that satisfy $A_t \geq \langle p_t, q_t \rangle$ for all $t \in \{1,\dots,T\}$.

The following proposition then states the wanted result; i.e., for a given data set $S = \{p_t, q_t\}_{t=1,\dots,T}$, finding a solution (in terms of the binary variables $x_{s,t}$) for this CS.I program is equivalent to S satisfying GARP.

Proposition 1. *Let $S = \{p_t, q_t\}_{t=1,\dots,T}$ be a set of observations. The following statements are equivalent:*

(i) S satisfies CS.I;

(ii) S satisfies GARP.

Proof. Suppose there exists a solution for program CS.I. If $q_t\, R_o\, q_s$, which follows from $\langle p_t, q_t \rangle \geq \langle p_t, q_s \rangle$, then condition (i) of CS.I implies that $x_{t,s} = 1$. Next, if $q_t\, R\, q_v$, which follows form $q_t\, R_o\, q_s$, $q_s\, R_o\, q_u$, $\dots$, and $q_r\, R_o\, q_v$, then $x_{t,v} = 1$ by condition (ii) of CS.I. Finally, if $q_t\, R\, q_s$, and thus $x_{t,s} = 1$, then the right hand side of condition (iii) of CS.I must be positive (i.e. $\langle p_s, q_t \rangle \geq \langle p_s, q_s \rangle$). Thus, we can conclude that S satisfies GARP.

Suppose then that S satisfies GARP and define $x_{s,t}$ as follows: $x_{s,t} = 1 \Leftrightarrow q_s\, R\, q_t$. Let us show that this is a solution for CS.I. By construction, condition (i) of CS.I is only restrictive if its left hand side is positive (i.e. $\langle p_t, q_t \rangle \geq \langle p_t, q_s \rangle$); but then $x_{s,t} = 1$ and thus rule (i) is satisfied. Condition (ii) of CS.I is met because of the transitivity of R. Finally, again by construction, condition (iii) is only restrictive if its left hand side is zero (i.e. $x_{t,s} = 1$); but then rule (iii) of GARP implies that the right hand side is also non-negative. This shows that we obtained a solution for CS.I. $\square$

6

Theorem 1 together with Proposition 1 gives us three distinct ways to verify whether a given data set $S = \{p_t, q_t\}_{t=1,\ldots,T}$ is rationalizable: (i) via the definition of GARP, (ii) via the Afriat inequalities and (iii) using CS.1. In what follows, we will compare the efficiency of these different methods for testing whether a given data set is rationalizable for the models in subsequent sections. Therefore, it is important to discuss them a bit more in detail.

The first method was originally suggested by Varian (1982), and therefore we call it the VARIAN-method. The method consists of three steps, which comply with the three conditions in Definition 2 of GARP. The first step constructs the relation R_0 from the data set $S = \{p_t, q_t\}_{t=1,\ldots,T}$. In particular $q_t R_0 q_s$ if and only if $\langle p_t, q_t \rangle \geq \langle p_t q_s \rangle$. A second step computes the transitive closure of R_0, i.e. the relation R. Varian (1982) suggests using Warshall's algorithm (Warshall, 1962), which is an efficient algorithm for computing transitive closures. The third step verifies if for all $q_s R q_t$ it is the case that $\langle p_t, q_t \rangle \leq \langle p_t, q_s \rangle$. If this is the case, the data set satisfies GARP and is, therefore, rationalizable. Due to its efficiency, the VARIAN-method is very popular in applied work. It will also serve as a valuable tool for developing quick heuristics when confronted with more elaborated models (see Section 3).

The second method verifies the rationalizability conditions by testing feasibility of the corresponding Afriat inequalities (i.e. condition (iii) of Theorem 1). These inequalities are linear in the unknowns ϕ_i and λ_i ($i \in \{1, \ldots, T\}$) which implies that their feasibility can be verified using elementary linear programming methods. We refer to Afriat (1967) and Diewert (1973) for discussions of this method. We call it the AFRIAT-method. A big advantage of this method is that it provides not only an efficient way to verify the rationalizability conditions but also, via the computed values of ϕ_i and λ_i, an estimate for the associated utility levels.

The third method verifies the rationalizability conditions via the conditions in CS.1. These conditions are linear in the unknown binary variables $x_{s,t}$. Therefore feasibility can be verified by standard integer programming (IP) methods (branch and bound, cutting plane, etc.). We refer to this method as the IP-method. Compared to the other methods, it is very inefficient and should not be recommended for applied work for the model developed in this section. However, in contrast to the other two methods, the IP method will be very useful in combination with the other restrictions of general equilibrium models. This will be discussed in the next section.

3 Pure exchange economies with assignable incomes

As in the previous section, we assume that there are J goods and T observations, but now we consider a pure exchange economy with N individuals. Each individual is endowed with a well-behaved utility function. The collection of these utility functions is denoted by $\{u^i\}_{i=1,\ldots,N}$. In each period t, we endow each individual i with an income I_t^i. Aggregate endowments in period t are given by a J-dimensional vector ε_t. The following concepts will be used throughout the paper.

Definition 3. *For given $\{u^i\}_{i=1,\ldots,N}$ and $\{\varepsilon_t\}_{t=1,\ldots,T}$, we define*

(i) $\{q_t^i\}_{t=1,...,T;i=1,...N}$ *is a feasible allocation if* $q_t^i \in \mathbb{R}_+^J$ *and* $\sum_{i=1}^{N} q_t^i = \varepsilon_t$;

(ii) $\{p_t, q_t^i\}_{t=1,...,T;i=1,...,N}$ *is a competitive equilibrium if* $\{q_t^i\}_{t=1,...,T;i=1,...N}$ *is a feasible allocation,* $q_t^i \in \underset{\langle p_t, q \rangle \leq \langle p_t, q_t^i \rangle}{\arg\max} u^i(q)$ *and* $p_t \in \mathbb{R}_{++}^J$. *The prices* p_t *are called the equilibrium prices.*

In words, condition (i) states the market clearing condition for each observation t; i.e., the quantities q_t^i allocated to the individuals i must add up to the aggregate endowment ε_t. Such an allocation represents a competitive equilibrium if, for given prices p_t, each q_t^i maximizes the utility of individual i; see condition (ii).

Brown and Matzkin (1996) start from observations on the set of equilibrium prices $\{p_t\}_{t=1,...,T}$, the set of aggregate endowments $\{\varepsilon_t\}_{t=1,...,T}$ and a set of individual incomes $\{I_t^i\}_{t=1,...,T;i=1,...,N}$. The following definition extends the earlier rationalizability concept to this general equilibrium context.

Definition 4. *A data set* $S = \{p_t, I_t^i, \varepsilon_t\}_{t=1,...,T;i=1,...,N}$ *is rationalizable if there exist well-behaved utility functions* $\{u^i\}_{i=1,...,N}$ *and a feasible allocation* $\{q_t^i\}_{t=1,...,T;i=1,...,N}$ *such that for all* $t = 1,...,T, i = 1,...N$:

(i) $\langle p_t, q_t^i \rangle = I_t^i$;

(ii) $\{p_t, q_t^i\}_{t=1,...,T;i=1,...,N}$ *is a competitive equilibrium.*

Compared to Definition 1, this definition accounts for N individuals instead of one. Condition (i) imposes the budget constraints implied by the observed individual incomes. Next, condition (ii) gives the corresponding competitive equilibrium requirement.

The next theorem recaptures the main result of Brown and Matzkin.

Theorem 2 (Brown and Matzkin). *The data set* $\left\{p_t, I_t^i, \varepsilon_t\right\}_{t=1,...,T;i=1,...,N}$ *is rationalizable if and only if for all* $t = 1,...T, i = 1,...,N$ *there exist numbers* $\phi_t^i \geq 0$, $\lambda_t^i > 0$ *and vectors* $q_t^i \in \mathbb{R}_+^J$ *such that:*

$$(3.1) \qquad\qquad \sum_{i=1}^{N} q_t^i = \varepsilon_t;$$

$$(3.2) \qquad\qquad \langle p_t, q_t^i \rangle = I_t^i;$$

$$(3.3) \qquad\qquad \phi_t^i \leq \phi_s^i + \lambda_s^i (\langle p_s, q_t^i \rangle - I_s^i).$$

Requirements (3.1) and (3.2) have been discussed before; requirement (3.3) captures the corresponding individual Afriat inequalities introduced in Theorem 1.

As indicated above, a main motivation for the IP approach developed in this paper is that verifying the restrictions in Theorem 2 is an NP-complete problem. Let us start by giving the suitable decision problem. For any numbers N, J and $T \in \mathbb{N}$ and a data set $S =$

$\{p_t, I_t^i, \varepsilon_t\}_{t=1,\ldots,T, i=1,\ldots,T}$, the problem RATIONALIZABILITY asks whether there exist vectors $\{q_t^i\}_{t=1,\ldots,T; i=1,\ldots,N}$ such that conditions (3.1), (3.2) and (3.3) of Theorem 2 are satisfied. The next theorem provides a formal statement of our result. We refer to the Appendix for the proof.

Theorem 3. *The decision problem* RATIONALIZABILITY *is* **NP***-complete.*

Thus, one should not focus on constructing polynomial time algorithms for testing the conditions in Theorem 2. By contrast, the result suggests that a more fruitful avenue consists in developing easy to implement and versatile non-polynomial time algorithms. Further, if these algorithms become computationally demanding (e.g., in the case of large data sets), it suggests using heuristics that give a quick (but possibly inconclusive) answer to whether or not the data satisfy the general equilibrium restrictions. This is exactly the approach that we follow in this paper. Moreover, it should be noted that for more specific cases, the rationalizability problem may allow for polynomial time verification.

The starting point of our IP method is given by the result in Proposition 1, which allows us to reformulate Theorem 2 by using the program CS.I.

Proposition 2. *The data set* $\left\{p_t, I_t^i, \varepsilon_t\right\}_{t=1,\ldots,T; i=1,\ldots,N}$ *is rationalizable if and only if for all* $t = 1,\ldots,T, i = 1,\ldots,N$ *there exist vectors* $q_t^i \in \mathbb{R}_+^J$ *such that:*

$$(3.4) \qquad\qquad \sum_{i=1}^{N} q_t^i = \varepsilon_t;$$

$$(3.5) \qquad\qquad \langle p_t, q_t^i \rangle = I_t^i;$$

$$(3.6) \qquad\qquad \forall i : \{p_t, q_t^i\} \text{ satisfies CS.I.}$$

Of course, in light of Proposition 1, we could have replaced condition (3.6) by the requirement that for all i: $\{p_t, q_t^i\}_{t=1,\ldots,T}$ satisfies GARP.

It would appear that the three methods for solving the rationalizability question that we discussed in the previous section (i.e. the VARIAN-, AFRIAT- and IP-method) can still be applied in this setting. However, this is not true. First of all, the VARIAN-method is no longer feasible. Indeed, as we no longer observe the quantities q_t^i, we are no longer able to construct the revealed preference relation R_o, which is needed in order to use Warshall's algorithm. Second, concerning the AFRIAT-method, observe that the Afriat inequalities in requirement (3.3) are quadratic since we have no observations on either λ_s^i or q_t^i. This turns the AFRIAT-method into verifying the feasibility for a set of quadratic inequalities, a problem that could be solved by using techniques for solving linear programs with quadratic constraints (LPQC).

The only approach that remains intact is the IP-method. Indeed, the inequalities in CS.I remains linear even if the quantities q_t^i are unobserved. Further, as is well known, IP problems can be solved more efficiently than LPQC problems.[6] This makes the IP-method the most preferred solution method for the model in this section.

---

[6]By adding for any binary variable x, the constraints $0 \le x \le 1$ and $x^2 = x$ we can easily convert any IP problem into a corresponding LPQC problem.

From elementary IP theory, we know that we can always verify in finite time whether a given IP problem is feasible. However, it is well-known that solving an IP problem may become computationally hard if the number of binary variables gets large. For such large problems one can always build further on Proposition 2 to derive heuristics that quickly (but possibly inconclusively) answer whether or not the data at hand satisfy the IP restrictions. For example, such an easy-to-implement heuristic uses elementary linear programming methods to verify whether there exists consumption bundles q_t^i that solve the linear equalities (3.4) and (3.5) and the linear relaxation of CS.I. More refined heuristics can explicitly incorporate additional restrictions of the CS.I program. In the next section we present some heuristics that apply to a variant of the model in this section.

4 Pure exchange economies without fully assignable incomes

An important restriction on the model in the previous section is the requirement that all individual incomes are observed. In reality, data sets often only capture (at best) partial information on the individual incomes. For example, lower income bounds can be defined on the basis of minimum income regulations (e.g., minimum wages) or because only labor income (and not capital income) is observed. In an extreme scenario, there is no information at all on the income distribution; e.g., this is the usual assumption in collective consumption models. In this section, we investigate the testable implications of an exchange model with less than perfect information on the individual incomes (i.e., with only information on bounds of the individual incomes). We also provide some quick heuristics that provide sufficient conditions for rationalizability.

We first introduce some additional notation. The real valued variables $h_t^i, l_t^i \in \mathbb{R}_+$ denote, respectively, upper and lower bounds on the income of individual i in period t. The associated notion of rationalizability is defined as follows:

Definition 5. *A data set $\{p_t, h_t^i, l_t^i, \varepsilon_t\}_{t=1,\ldots,T; i=1,\ldots,N}$ is rationalizable if there exist well-behaved utility functions $\{u^i\}_{i=1,\ldots,N}$ and a feasible allocation $\{q_t^i\}_{t=1,\ldots,T; i=1,\ldots,N}$ such that for all $t = 1,\ldots,T, i = 1,\ldots,N$:*

(i) $l_t^i \leq \langle p_t, q_t^i \rangle \leq h_t^i$;

(ii) $\{p_t, q_t^i\}_{t=1,\ldots,T; i=1,\ldots,N}$ is a competitive equilibrium.

The interpretation is similar to the one of Definition 4. The main difference is that we no longer fully observe the individual incomes; and, thus, condition (i) incorporates the income bounds information that is available. The following proposition provides a straightforward extension of Proposition 2.

Proposition 3. *A data set $\{p_t, h_t^i, l_t^i, \varepsilon_t\}_{t=1,\ldots,T; i=1,\ldots,N}$ is rationalizable if and only if for all $t = 1,\ldots,T, i = 1,\ldots,N$ there exist vectors $q_t^i \in \mathbb{R}_+^J$ such that:*

$$(4.1) \qquad \sum_{i=1}^{N} q_t^i = \varepsilon_t;$$

$$(4.2) \qquad l_t^i \leq \langle p_t, q_t^i \rangle \leq h_t^i;$$

$$(4.3) \qquad \forall i : \{p_t, q_t^i\}_{t=1,\dots,T} \text{ satisfies CS.I.}$$

Example 1 below shows that the condition in Proposition 3 can be rejected as soon as we have two goods and two observations with strictly positive lower bounds on the income of one individual.[7] Given that we do not assume fully assignable incomes, this conclusion generalizes the result of Brown and Matzkin (1996); these authors have shown refutability of the condition in Theorem 2 for two observations and two goods in the case of fully assignable incomes.

Example 1. *For all T, N and J with $T, J \geq 2$, and real numbers $l_t^i > 0$, $l_v^i > 0$, there exists a data set $\{p_t, h_t^i, l_t^i, \varepsilon_t\}_{t=1,\dots,T; i=1,\dots,N}$ that is not rationalizable for any feasible allocation $\{q_t^i\}_{t=1,\dots,T; i=1,\dots,N}$ with $\langle p_t, q_t^i \rangle \geq l_t^i$ and $\langle p_v, q_v^i \rangle \geq l_v^i$.*
It suffices to consider $T = J = 2$. Let

$$p_t = \left(\frac{l_t^i}{10}, \frac{l_t^i}{100} \right),$$

$$p_v = \left(\frac{l_v^i}{100}, \frac{l_v^i}{10} \right),$$

$$\varepsilon_t = (10, 1),$$

$$\varepsilon_v = (1, 10).$$

We see that $\langle p_t, \varepsilon_t \rangle > l_t^i > \langle p_t, \varepsilon_v \rangle$ and that $\langle p_v, \varepsilon_v \rangle > l_v^i > \langle p_v, \varepsilon_t \rangle$. If the data set is rationalizable, we must have that there exists $\{q_t^i\}_{t=1,\dots,T; i=1,\dots,N}$ such that:

$$\langle p_t, q_t^i \rangle \geq l_t^i,$$

$$\langle p_v, q_v^i \rangle \geq l_v^i,$$

$$\langle p_t, \varepsilon_v \rangle \geq \langle p_t, q_v^i \rangle,$$

$$\langle p_v, \varepsilon_t \rangle \geq \langle p_v, q_t^i \rangle.$$

Clearly, this implies a violation of the condition in Proposition 3; the GARP condition for individual i cannot be satisfied.

As discussed in the previous section, verifying whether the IP program in Proposition 3 is feasible can be computationally very demanding. Therefore, we next introduce sufficient conditions for rationalizability that are generally easier to implement than the IP characterization

---

[7]Example 1 only uses information on lower income bounds for illustrative purposes. Clearly, examples with information on upper income bounds can proceed in a readily analogous manner.

in Proposition 3; their computational complexity depends on the number of observations and goods. These conditions will also prove useful to design easy-to-implement heuristics.

Let us start by giving some intuition. Assume that we divide the set of agents $\{1,\ldots,N\}$ into groups of more or less equal size (details will be provided in the formal treatment below). Then, we partition the set of goods and assign to each group of individuals at random an element of this partition. Finally, we divide the economy wide endowments of the goods assigned to a certain group equally among its members. Our result then determines the conditions for which this kind of allocation provides a solution to the rationalizability problem.

In order to formalize our intuition, we start from a data set $S = \{p_t, l_t^i, h_t^i, \varepsilon_t\}_{t=1,\ldots,T}$ and consider a partition $\mathscr{P}$ for the set of goods $\{1,\ldots,J\}$. Each element of this partition determines the set of goods for a certain group of individuals. For each such element $P \in \mathscr{P}$, we define the J-dimensional vectors $q_t^P = ((q_t^P)_1,\ldots,(q_t^P)_J)$ as follows:

$$(q_t^P)_j = \begin{cases} (\varepsilon_t)_j \text{ if } j \in P \\ 0 \text{ if } j \notin P \end{cases}$$

In words, the partition $\mathscr{P}$ allocates every endowment ε_t such that each quantity $(\varepsilon_t)_j$ is fully assigned to exactly one vector q_t^P.

The next result determines sufficient conditions for which the data set S is rationalizable given the number of individuals $N \geq |\mathscr{P}|$.[8]

Proposition 4. *Consider a data set $S = \{p_t, h_t^i, l_t^i, \varepsilon_t\}_{t=1,\ldots,T; i=1,\ldots,N}$ and assume that $\{1,\ldots,J\}$ has a partition $\mathscr{P}$ such that for all $P \in \{\mathscr{P}\}$, $\{p_t, q_t^P\}_{t=1,\ldots,T}$ satisfies* GARP.

Then, the data set is rationalizable if $N \geq |\mathscr{P}|$, and for all $P \in \mathscr{P}, t = 1,\ldots,T, i = 1,\ldots,N$:

$$\max\{\eta - 1, 1\} h_t^i \geq \langle p_t, q_t^P \rangle \geq \eta l_t^i,$$

where

$$\eta = \left\lceil \frac{N}{|P|} \right\rceil$$

Recall that $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

Proof. Consider the set of individuals $\{1,\ldots N\}$ and let $\gamma : \{1,\ldots,N\} \to \mathscr{P}$ be a surjection from the set of individuals to the partition. For $P \in \mathscr{P}$, we denote by $|\gamma^{-1}(P)|$ the cardinality of the set $\{\gamma(i) = P\}$. By definition of η, we may assume that the surjection γ is such that $\eta - 1 \leq |\gamma^{-1}(P)| \leq \eta$ for all $P \in \mathscr{P}$.

Let us then define the following allocation

$$q_t^i = \frac{q_t^P}{|\gamma^{-1}(\gamma(i))|}.$$

---

[8]Recall that $|\mathscr{P}|$ denotes the cardinality of $\mathscr{P}$.

This is a feasible allocation, since for all $t = 1, \ldots, T$:

$$\sum_{i=1}^{N} q_t^i = \sum_{P \in \mathscr{P}} \sum_{\{i, \gamma(i) = P\}} \frac{q_t^P}{|\gamma^{-1}(\gamma(i))|} = \sum_{P \in \mathscr{P}} q_t^P = \varepsilon_t.$$

Next, the conditions of the proposition and $\eta - 1 \le |\gamma^{-1}(P)| \le \eta$, imply that $h_t^i \ge \langle p_t, q_t^i \rangle \ge l_t^i$. Finally given that $\{p_t, q_t^P\}_{t=1,\ldots,T}$ satisfies GARP, we have that the $\{p_t, q_t^i\}_{t=1,\ldots,T}$ satisfies GARP for all i. As such we obtain that there must exist a solution for Proposition 3. $\square$

For a given data set $\{p_t, \varepsilon_t\}_{t=1,\ldots,T}$, the collection of partitions $\mathscr{P}$ can be constructed in finite time. Moreover, the GARP condition in Proposition 4 can be checked very efficiently by using the VARIAN-method discussed at the end of Section 2. This implies that, although the procedure in Proposition 4 may be exponential in the number of goods, J, it is polynomial in the number of individuals and the number of observations.

The procedure in Proposition 4 could easily be adapted to allow for groups of different sizes or for partitions that differ across observations. For example, consider a setting where we allow a group of individuals either to consume a subset of goods or nothing at all. Obviously, for observations where consumption is zero, the lower bound on income also has to be zero, hence for simplicity we assume that $l_t^i = 0$ and $h_t^i = \langle p_t, \varepsilon_t \rangle$ for all $t \in \{1, \ldots, T\}$ and $i \in \{1, \ldots, N\}$. This complies with having no information at all concerning the individual incomes. It also implies that condition (4.2) is always satisfied, so that we can ignore it for the moment. Now consider a partition $\mathscr{P}$ for the set $\{1, \ldots, J\} \times \{1, \ldots, N\}$ and in an analogue fashion we define:

$$(q_t^P)_j = \begin{cases} (\varepsilon_t)_j \text{ if } (j, t) \in P \\ 0 \text{ if } (j, t) \notin P \end{cases}$$

The following heuristic is a straightforward variation on Proposition 4.

Proposition 5. *Consider a data set $\{p_t, \varepsilon_t\}_{t=1,\ldots,T}$ and a partition $\mathscr{P}$ of $\{1, \ldots, J\} \times \{1, \ldots, T\}$. Assume that for all $P \in \mathscr{P}$, the set $\{p_t, q_t^P\}_{t=1,\ldots,T}$ satisfies GARP. Then, the data set is rationalizable for all exchange economies with a number of individuals $N \ge |\mathscr{P}|$.*

Proof. The proof is analogous to the proof of Proposition 4 and is left to the reader. $\square$

We note that the procedure from Proposition 4 only uses partitions for the set $\{1, \ldots, J\}$, which imposes extra structure as compared to the partitions in Proposition 5. As such, checking all the GARP conditions for Proposition 5 will be computationally more demanding than checking these for Proposition 4; in other words, the procedure in Proposition 5 may be exponential in both the number of goods J and the number of observations T. However, we point out that the result in Proposition 5 can provide a useful basis for designing even easier heuristics, which avoid specifying all partitions $\mathscr{P}$ of $\{1, \ldots, J\} \times \{1, \ldots, T\}$. To illustrate this last point, the next result recaptures existing results in the literature; in particular, it reproduces Theorem 3 of Snyder (2004).

Heuristic 1. *Let $\{p_t, \varepsilon_t\}_{t=1\ldots,T}$ be given.*

1. If $N \geq J$ then $\{p_t, \varepsilon_t\}$ is rationalizable.

2. If $N \geq T$, then $\{p_t, \varepsilon_t\}$ is rationalizable.

Proof. It is sufficient to show that for any case we can construct a partition $\mathscr{P}$ such that $\{p_t, q_t^P\}_{t=1,\dots,T}$ satisfies GARP for all $P \in \mathscr{P}$. For case 1, we consider the partition $\{\{j\} \times \{1,\dots,T\}|j = 1,\dots,J\}$. For case 2, we consider the partition $\{\{1,\dots,J\} \times \{t\}|t = 1,\dots,T\}$. It is straightforward to show that all partitions satisfy the requirements of Proposition 5. $\quad\square$

Heuristic 1 implies that, if we assume that only prices and aggregate endowments are observable, then the number of individuals must be smaller than both the number of observations and the number of goods (i.e., $N < T$ and $N < J$) if we want to reject rationalizability. Clearly, this implication puts serious limitations on the empirical usefulness of equilibrium models such as, e.g., the one in Proposition 5. Still, at this point it is worth to recall once more the possibility to analyze small group behavior on the basis of general equilibrium theory; example applications concern multi-person household consumption, as repeatedly indicated before, but also trade between a limited number of countries. For such small group settings, we easily obtain $N < T$ and $N < J$. Therefore, all the heuristics developed in this section can also be included into the list of already existing heuristics for models of small group behavior (e.g. Deb, 2008; Talla Nobibon et al., 2008).

The following result allows us to quickly find a rationalization of data sets that do not satisfy the conditions of Heuristic 1 (i.e., $N < T$ and $N < J$).

Heuristic 2. *For any $k \in \mathbb{N}$: if $N \geq \dfrac{T}{k}$ and any subset of k observations of $\{p_t, \varepsilon_t\}_{t=1\dots,T}$ satisfies* GARP, *then $\{p_t, \varepsilon_t\}$ is rationalizable.*

Proof. Take the following partition of $\mathscr{V}: \{\{1,\dots,k\}, \{k+1,\dots,2k\},\dots,\{[\dfrac{T}{k}]k,\dots,T\}\}$. Clearly, each element of this partition contains at most k elements and as such we have that the corresponding sets of prices p_t and endowments ε_t satisfy GARP. A similar argument as in Proposition 5 then finishes the proof. $\quad\square$

In graph theoretical terms, this heuristic boils down to finding paths of a given length. As such, practical applications of the heuristic can use well-established and efficient algorithms in the literature on graph theory, such as Dijkstra's algorithm.

5 Economies with public goods and consumption externalities

The IP characterizations in Propositions 2 and 3 allow for fairly straightforward extensions that imply refinements of the basic general equilibrium model. In this section, we illustrate the versatility of the proposed IP approach by characterizing general equilibrium models that

account for public consumption and externalities. The next section provides some additional extensions.

Two preliminary remarks are in order. Firstly, we assume Pareto optimal allocations. Admittedly, in the presence of public goods and/or externalities, a feasible allocation may be Pareto suboptimal. As a matter of fact, this makes it all the more interesting to empirically test whether the Pareto efficiency assumption effectively holds; see also Snyder (1999) for discussion on this Pareto efficiency benchmark in models with public goods and externalities. The IP characterizations described below allow for conducting such efficiency tests. Secondly, we will assume that the public goods and the externalities are all positively evaluated by the individuals in the economy. Indeed, Snyder (1999) argued that we need this assumption to obtain refutable models.

To structure our following exposition, we will focus on two extreme cases: the case with all consumption public; and the case without public consumption, and with all private consumption characterized by externalities. Of course, intermediate cases can be considered in practical applications. For example, the same good can simultaneously be a private good with no externality (e.g. a tree in a private garden), an externality generating good (e.g. a tree casting a shadow on a neighboring garden) and a public good (e.g. a tree in a park). Because we believe the analogy is fairly easy, we will not explicitly consider these intermediate cases.

We first consider the case in which all consumption is public. In this case we assume that the aggregate endowment of public goods ε can be converted to the bundle of public consumption goods q according to a constant returns to scale production technology. This implies that the equilibrium price vector, p, must satisfy the restriction

$$\langle p, q - \varepsilon \rangle = 0;$$

see, for example, Theorem 6 of Varian (1984).

Next, because all consumption is public, the utility function of each individual i can be written as $u^i(q)$. A standard result (see Bergstrom, 1976) is that any Pareto efficient allocation of publicly consumed goods can be implemented by a set of personalized Lindahl prices p_t^i ($t = 1, \ldots, T$ and $i = 1, \ldots, N$), which must add up to the equilibrium price vector p_t, i.e.

$$(5.1) \qquad p_t = \sum_{i=1}^{N} p_t^i.$$

Given the use of Lindahl prices, we will refer to Lindahl equilibrium instead of competitive equilibrium in the next rationalizability definitions. For compactness, we will not include a formal definition of the Lindahl equilibrium concept; but this definition is straightforwardly similar to Definition 3 of the competitive equilibrium concept, with Lindahl prices p_t^i instead of observed prices p_t. In what follows, we say that $\{p_t^i\}_{t=1,\ldots,T; i=1,\ldots,N}$ is a personalized price system if the vectors $p_t^i \in \mathbb{R}_{++}^J$ satisfy the adding up condition (5.1).

Thus, we get the following rationalizability concept, which straightforwardly extends the definitions in Section 3.

Definition 6. *A data set* $\{p_t, h_t^i, l_t^i, \varepsilon_t, q_t\}_{t=1,\dots,T; i=1,\dots N}$ *is rationalizable if there exist well-behaved utility functions* $\{u^i\}_{i=1,\dots,N}$ *and a personalized price system* $\{p_t^i\}_{t=1,\dots,T; i=1,\dots,N}$ *such that for all* $t = 1,\dots,T, i = 1,\dots,N$:

(i) $l_t^i \leq \langle p_t^i, q_t \rangle \leq h_t^i$;

(ii) $\langle p_t, q_t - \varepsilon_t \rangle = 0$, *and*

(iii) $\{p_t^i, q_t\}_{t=1,\dots,T; i=1,\dots,N}$ *is a Lindahl equilibrium.*

Once more, rationalizability requires consistency with GARP conditions (or Afriat inequalities); but now these GARP conditions are expressed in terms of personalized prices. Analogous reasoning as before obtains the following result.

Proposition 6. *A data set* $\{p_t, h_t^i, l_t^i, \varepsilon_t, q_t\}_{t=1,\dots,T; i=1,\dots,N}$ *is rationalizable if and only if for all* $t = 1,\dots,T, i = 1,\dots,N$ *there exist vectors* $p_t^i \in \mathbb{R}_+^J$ *such that:*

(i) $p_t = \sum_{i=1}^N p_t^i$;

(ii) $\langle p_t, q_t - \varepsilon_t \rangle = 0$;

(iii) $l_t^i \leq \langle p_t^i, q_t \rangle \leq h_t^i$;

(iv) $\forall i : \{p_t^i, q_t\}_{t=1,\dots,T}$ *satisfies* CS.I.

The interpretation parallels to the one of Proposition 3. Like before, we conclude that elementary IP methods can verify in finite time whether a given data set is rationalizable.[9]

Let us then consider the case in which there is no public consumption and all private consumption is characterized by externalities; the utility function of each individual is specified as $u\left(q^1, \dots, q^N\right)$. In this case, the treatment of privately consumed quantities is formally similar to the above treatment of publicly consumed quantities; i.e., Pareto efficiency requires Lindahl pricing. Specifically, for each quantity vector q_t^j, the corresponding personalized prices satisfy

$$p_t = \sum_{i=1}^N p_t^{i,j},$$

with $p_t^{i,j}$ representing the willingness-to-pay of individual i for the consumption of individual j.

To simplify the notation, we let $Q_t = (q_t^1, \dots, q_t^N)$ and $P_t^i = (p_t^{i,1}, \dots, p_t^{i,N})$. In what follows, a feasible allocation $\{Q_t\}_{t=1,\dots,T}$ satisfies $\sum_{i=1}^N q_t^i = q_t$ with $q_t^i \in \mathbb{R}_+^J$, and a personalized price system $\{P_t^i\}_{t=1,\dots,T; i=1,\dots,N}$ satisfies $p_t = \sum_{i=1}^N p_t^{i,j}$ with $p_t^{i,j} \in \mathbb{R}_+^J$. Using this, we obtain the following rationalizability definition when again assuming a constant returns to scale production technology.

---

[9]It can easily be shown that the Afriat inequalities corresponding to condition (iv) of Proposition 6 are given by a set of quadratic inequalities. Hence, the IP-method is again deemed more efficient then the AFRIAT-method.

Definition 7. *A data set $\{p_t, h_t^i, l_t^i, \varepsilon_t, q_t\}_{t=1,\ldots,T; i=1,\ldots N}$ is rationalizable if there exist well-behaved utility functions $\{u^i\}_{i=1,\ldots,N}$, a feasible allocation $\{Q_t\}_{t=1,\ldots,T}$ and a personalized price system $\{P_t^i\}_{t=1,\ldots,T; i=1,\ldots,N}$ such that for all $t = 1,\ldots,T, i = 1,\ldots,N$:*

(i) $l_t^i \le \langle P_t^i, Q_t \rangle \le h_t^i$;

(ii) $\langle p_t, q_t - \varepsilon_t \rangle = 0$ and

(iii) $\{P_t^i, Q_t\}_{t=1,\ldots,T; i=1,\ldots,N}$ is a Lindahl equilibrium.

The next result gives the corresponding rationalizability characterization; the argument proceeds similarly as before.

Proposition 7. *A data set $\{p_t, h_t^i, l_t^i, \varepsilon_t, q_t\}_{t=1,\ldots,T; i=1,\ldots,N}$ is rationalizable if and only if for all $t = 1,\ldots,T, i,j = 1,\ldots,N$ there exist vectors $p_t^{i,j} \in \mathbb{R}_+^J$ and $q_t^i \in \mathbb{R}_+^J$ such that, for $Q_t = (q_t^1,\ldots,q_t^N)$ and $P_t^i = (p_t^{i,1},\ldots,p_t^{i,N})$:*

(i) $\varepsilon_t = \sum_{i=1}^N q_t^i$ and $p_t = \sum_{i=1}^N p_t^i$;

(ii) $\langle p_t, q_t - \varepsilon_t \rangle = 0$,

(iii) $l_t^i \le \langle P_t^i, Q_t \rangle \le h_t^i$;

(iv) $\forall i : \{P_t^i, Q_t\}_{t=1,\ldots,T}$ satisfies CS.I.

Importantly, the CS.I program in this result is no longer linear in terms of unknowns, which contrasts with earlier CS.I programs. Specifically, the program requires that, for all $i = 1,\ldots,N$, there exist binary variables $\{x_{s,t}^i\}_{s,t=1\ldots,T; i=1,\ldots,N}$ such that for all $t,s,v = 1,\ldots,T$:

$$\langle P_t^i, Q_t \rangle - \langle P_t^i, Q_s \rangle < x_{t,s}^i A_t^i;$$
$$x_{t,s}^i + x_{s,v}^i \le 1 + x_{t,v}^i;$$
$$(x_{t,s}^i - 1)A_s^i \le \langle P_s^i, Q_t \rangle - \langle P_s^i, Q_s \rangle;$$

with $A_t^i \ge \langle p_t, Q_t \rangle$. This clearly implies non-linearity in prices P_t^i and quantities Q_t.[10] In order to restore linearity, we define $y_{t,s} = \langle p_t, q_s \rangle$ and $\alpha_{t,s}^i = \dfrac{\langle P_t^i, Q_s \rangle}{y_{t,s}}$; $y_{t,s}$ represents the aggregate value of consumption bundle q_s under prices p_t, and $\alpha_{t,s}^i$ represents the share of this value allocated to individual i. Consider then the following set of restrictions for each $i = 1,\ldots,N$:

---

[10]In this case, the Afriat inequalities are determined by a set of cubic inequalities.

Program (CS.II).

$$\sum_{i=1}^{n}\alpha_{t,s}^{i} = 1 \text{ and } l_{t}^{i} \le \alpha_{t,t}^{i} y_{t,t} \le h_{t}^{i} \quad (t,s=1,\ldots,T);$$

$$\alpha_{t,t}^{i} y_{t,t} - \alpha_{t,s}^{i} y_{t,s} < x_{t,s} A_{t}^{i} \qquad (t,s=1,\ldots,T; i=1,\ldots,N);$$

$$x_{t,s}^{i} + x_{s,v}^{i} \le 1 + x_{t,v}^{i} \qquad (t,s,v=1,\ldots,T; i=1,\ldots,N);$$

$$(x_{t,s}^{i} - 1)A_{s}^{i} \le \alpha_{s,t}^{i} y_{s,t} - \alpha_{s,s}^{i} y_{s,s} \qquad (t,s=1,\ldots,T; i=1,\ldots,N).$$

Where the constants A_t are given real numbers that satisfy $A_t \ge y_{t,t}$.

Clearly, these restrictions are linear, and feasibility (i.e., existence of $\{\alpha_{t,s}^{i}\}_{t,s=1,\ldots T; i=1,\ldots,N}$ and $\{x_{t,s}^{i}\}_{t,s=1,\ldots,T; i=1,\ldots,N}$ that satisfy CS.II) requires that the feasible region of the associated mixed integer linear program (MILP) is nonempty. Thus, we can verify feasibility in finite time by using elementary IP methods. Additional restrictions on Q_t and/or P_t^i can be included through (linear) restrictions on the $\alpha_{t,s}^{i}$. Feasibility of CS.II for all $i \in N$ constitutes a necessary (but not sufficient) condition for rationalizability in terms of Definition 7. Specifically, we obtain CS.II by substituting $y_{t,s}$ and $\alpha_{t,s}^{i}$ in the CS.I programs that apply to the individuals i (in Proposition 8); i.e., we use $\alpha_{t,s}^{i} y_{t,s}$ instead of $\langle P_t^i, Q_s \rangle$. This leads to the following conclusion.

Proposition 8. *A data set $\{p_t, h_t^i, l_t^i, \varepsilon_t\}_{t=1,\ldots,T; i=1,\ldots,N}$ is rationalizable only if it satisfies conditions (i) and (ii) of Proposition 7 and, for $y_{t,s} = \langle p_t, \varepsilon_s \rangle$, CS.II is a feasible program in terms of the variables $\alpha_{t,s}^{i} \ge 0$ and $x_{s,t}^{i} \in \{0,1\}$ for all $i \in N$.*

It is easily demonstrated that the data set of Example 1 in Section 4 does not satisfy CS.II together with conditions (i) and (ii) of Proposition 7, which implies the restrictions are non-vacuous. In fact, the conditions can be rejected even if there is no information on assignable incomes. For example, it can be verified the data set in Example 1 of Cherchye et al. (2007a) also rejects the conditions of Proposition 8.

6   Further extensions

Three further interesting extensions of the proposed IP characterizations are the following: (1) include more general production technologies in the general equilibrium model; (2) recover the structural model that underlies the observed equilibrium behavior; and (3) account for goodness-of-fit and power of the general equilibrium models in empirical analysis. These extensions are easily incorporated in the IP framework presented in the previous sections. For compactness, we abstract from a detailed formal discussion, but restrict to sketching the main arguments.

1. **Production**   General equilibrium settings often involve not only (utility maximizing) consumers but also (profit maximizing) producers. Therefore, it seems most useful to analyze the testable implications of general equilibrium models that include production. As indicated above, this can also be useful within the context of collective consumption models, as it paves the way for including household production in such settings.

We obtain IP characterizations for general equilibrium in production economies by combining our above results with those of Carvajal (2005). Specifically, under the technological assumptions of constant returns to scale and/or 'no free lunch', Carvajal derived refutable rationalizability conditions that only require data on aggregate production (i.e., without the observation of individual production levels). Interestingly, these testable conditions merely add linear production-type restrictions to the consumption-type (GARP) conditions considered before. As such, we can simply add them to our earlier IP characterizations (e.g., in Propositions 2 and 3) without interfering with the MILP structure of the resulting restrictions. See also Section 5, which considers constant returns to scale production technologies in the IP characterizations.

2. **Recoverability**   In many cases, testing whether a data set satisfies the restrictions of some behavioral (in casu general equilibrium) model only constitutes a first step in the empirical analysis. The next step wants to recover the structural model that underlies the observed behavior; in the context of general equilibrium models, such recovery focuses on reconstructing the individual preferences and individual quantity bundles from the observed equilibrium behavior. Chiapppori et al. (1999), for example, derived identifiability/recoverability results for general equilibrium models, which enable recovery by starting from some a priori (parametric) specification of the individual utility functions.

The IP characterizations developed above allow for adopting a so-called 'revealed preference' approach to address recoverability questions. The distinguishing feature of this approach is that it only uses the preference information that is revealed by the data; it does not require a (usually non-verifiable) prior specification for the utility functions. Typically, this revealed preference approach does not identify a unique specification of preferences from a finite set of data (which is different from the (parametric) approach in Chiapppori et al. (1999)); recovery aims at constructing inner and outer bounds on indifference curves, consumption bundles and production bundles. See, for example, Afriat (1967), Varian (1982) and Varian (2006) for detailed discussions of revealed preference recoverability; these authors consider the basic setting with a single consumer, and thus start from the rationalizability concept in Definition 1.

As for the general equilibrium models considered in this paper, recovery can proceed along the lines of Cherchye et al. (2007), who focused on collective consumption models (possibly characterized by public goods and externalities). To give a specific example, let us consider recovery of the individual quantities q_t^i on the basis of the IP characterizations in Propositions 2 and 3. We recall that checking the IP conditions boils down to verifying whether the corresponding MILP restrictions define a nonempty feasible region. Given this, we can recover a lower (resp. upper) bound by minimizing (resp. maximizing) q_t^i subject to the given MILP restrictions.

3. Empirical issues: goodness-of-fit and power   In view of practical applications, two important considerations concern (1) the goodness-of-fit of the general equilibrium model when data do not pass the ('sharp') condition for equilibrium behavior that is tested, and (2) the power of the nonparametric tests in terms of the probability of detecting non-equilibrium or 'irrational' behavior. Generally, a fair assessment of behavioral models should complement a goodness-of-fit analysis with a power analysis: favorable goodness-of-fit results, indicating few violations of the behavioral restrictions, have little meaning if the behavioral implications have low power, i.e., optimizing behavior can hardly be rejected. Conveniently, as we will argue, our IP approach allows for incorporating such considerations.

Firstly, Cherchye et al. (2008) propose two goodness-of-fit measures that can be calculated by starting from IP characterizations of collective consumption models.[11] The first measure is inspired by Varian (1990)'s idea to quantify goodness-of-fit in terms of optimization error; it can be interpreted as a measure for the economic significance of observed violations of optimizing behavior. The second measure is based on Varian (1985)'s idea to quantify goodness-of-fit in terms of measurement error, and can be interpreted as a measure for the statistical significance of observed violations of optimizing behavior. By using our IP characterizations of general equilibrium models, it is easy to introduce these measures into the current setting.

Secondly, Bronars (1987) presented a procedure that is specifically designed for detecting the probability of 'irrational' behavior for Afriat-type tests such as ours; see also Andreoni and Harbaugh (2006) for a recent review of power assessment tools. We can easily operationalize these power measures in the context of general equilibrium models on the basis of the IP characterizations presented above. We again refer to Cherchye et al. (2008) for more discussion; while these authors focus on IP characterizations of collective consumption models, their arguments readily extend to our IP characterizations of general equilibrium models.

Appendix: proof of Theorem 3

A *decision problem* can be described by an INSTANCE, the inputs, and a QUESTION, which has either "yes" or "no" as an answer. The general equilibrium rationalizability problem (which we will call RATIONALIZABILITY can be described as follows:

INSTANCE RATIONALIZABILITY: integers T, N and J, a data set $\{p_t, \varepsilon_t, I_t^i\}_{t=1,\dots,T; i=1,\dots,N}$ where the elements p_t are strict positive J-dimensional price vectors, the elements ε_t are non-negative J-dimensional endowment vectors and the numbers I_t^i are non-negative incomes.

QUESTION RATIONALIZABILITY: does there exist a set $\{q_t^i\}_{t=1,\dots,T; i=1,\dots,N}$ of positive J-dimensional consumption vectors that satisfies conditions (3.1), (3.2) and (3.3) of Theorem 2.

In what follows, we will use that condition (3.3) is equivalent to the condition that for all i: $\{p_t, q_t^i\}_{t=1,\dots,T}$ satisfies GARP (see Theorem 1).

A proof of NP-completeness proceeds in two steps: first, one has to demonstrate that the problem is in NP; and, second, one has to show that it is harder than any other NP problem.

---

[11]See also Kübler and Schmedders (2008), who discuss the notion of approximating equilibria, for a conceptually close approach within the specific context of (in casu dynamic stochastic) general equilibrium analysis.

The way to solve the second problem is to show that a known NP-complete problem is polynomial time reducible to the given problem. In particular, a problem P_1 is polynomial time reducible into a problem P_2 if (i) there exist a function g which maps every instance of P_1 into an instance of P_2 in such a way that a solution I_1 is a "yes" for P_1 if and only if $g(I_1)$ is a "yes" for P_2, and (ii) $g(I_1)$ can be constructed in polynomial time.

Given all this, to prove Theorem 3, we start by showing that RATIONALIZABILITY is in the class NP. Indeed, given a solution $\{q_t^i\}_{t=1,\dots,T;i=1,\dots,N}$ that satisfies the conditions in Theorem 2, conditions (3.1) and (3.2) are linear equalities and require only polynomial time to be verified. Also, condition (3.3) can be verified in polynomial time by using, for example, Varian's algorithm for verifying GARP (which uses Warshall's algorithm to compute the transitive closure of the revealed preference relation); see Varian (1982) for more discussion.

For the second part of the proof, we need to reduce a known NP-complete problem into an instance of RATIONALIZABILITY. To this end, we use the problem Monotone 3-SAT (M3SAT); see Garey and Johnson (1979) for a discussion of M3SAT.

INSTANCE M3SAT: A set of Boolean variables $\{x_1,\dots,x_n\}$ and a set of clauses $\{c_1,\dots,c_m\}$. Each clause is composed of exactly three distinct literals and a literal is either equal to a variable or its negation. The term monotone refers to the restriction that for all clauses, all literals in this clause are either negated or unnegated.

QUESTION M3SAT: Does there exist an assignment to the Boolean variables such that each clause contains at least one true literal? If the answer is "yes" we say that M3SAT is satisfiable.

We say that two clauses in an M3SAT instance are of *opposite signature* if one of the two clauses only contains literals that are unnegated and the other clause only contains literals that are negated.

Step 1: construction of the RATIONALIZABILITY instance

Consider an instance of M3SAT with a set of variables $\{x_1,\dots,x_n\}$ and a set of clauses $\{c_1,\dots,c_m\}$. First we create the set of individuals, the set of observations and the set of goods.

- For each variable $x_i \in \{x_1,\dots,x_n\}$ we create an individual i and these are the only individuals in our economy. This gives us n individuals (i.e. $N = n$).

- For each clause $c \in \{c_1,\dots,c_m\}$, we create two observations t_c and v_c and these are the only observations in the data set. This gives us $2m$ observations (i.e. $T = 2m$).

- For each clause c, we create a good denoted by $g(t_c, v_c)$. For every pair of clauses c and d of opposite signature, we create two goods $g(v_c, t_d)$ and $g(v_d, t_c)$ and these are the only goods in our economy. Let G be the set of goods. The cardinality of G is at most $m + m(m-1)$ (i.e. $J \le m + m(m-1)$).

Before we provide the values of the prices, aggregate endowments and income levels, we partition for each observation the set of goods, G, into five groups. For each clause c and the corresponding observations t_c and v_c, we define:

- $S(t_c) = \{g(t_c, v_c)\} = S(v_c)$,

- $S'(t_c) = \{g(t_d, v_d) \in G \mid d \neq c\} = S'(v_c)$,

- $O(t_c) = \{g(v_d, t_c) \in G \mid d \neq c\}$ and $O(v_c) = \{g(v_c, t_d) \in G \mid d \neq c\}$.

- $O^2(t_c) = \{g(v_d, t_e) \in G \mid g(v_d, t_c) \in O(t_c), g(v_d, t_e) \in O(v_d)\} - O(t_c)$ and
  $O^2(v_c) = \{g(v_e, t_d) \in G \mid g(v_c, t_d) \in O(v_c), g(v_e, t_d) \in O(t_d)\} - O(v_c)$,

- $E(t_c)$ contains all goods in G which are not accounted for in $S(t_c), S'(t_c), O(t_c)$ or $O^2(t_c)$ and $E(v_c)$ contains all goods in G which are not accounted for in $S(v_c), S'(v_c), O(v_c)$ or $O^2(v_c)$.

Clearly, the above sets have by construction an nonempty intersection and constitute for a given observation a corresponding partition of the set G. The sets $S(t_c), S(v_c), S'(t_c), S'(v_c), E(t_c)$ and $E(v_c)$ are self explanatory. A good $g(v_d, t_c)$ is in $O(t_c)$ and $O(v_c)$ if d and c are opposite clauses. A good $g(v_d, t_e)$ is in $O^2(t_c)$ if c and d are opposite clauses and at the same time e and d are also opposite clauses; to have an empty intersection with $O(t_c)$ we demand that e is not equal to c. A similar interpretation holds for a good $g(v_e, t_d)$ in $O^2(v_c)$.

For each observation we define the prices, p, and the endowments, ε, of the goods by using the set of goods that contains the given goods. These p and ε are given in Table 1 below (prices are between brackets).

Table 1: Endowments and prices

	$S(t_c)$	$S'(t_c)$	$O(t_c)$	$O^2(t_c)$	$E(t_c)$
t_c	$1\ (1)$	$6\ (\sigma^2)$	$1/3 - \sigma\ (\sigma^2)$	$1/3\ (\sigma^2)$	$3\ (\sigma^3)$

	$S(v_c)$	$S'(v_c)$	$O(v_c)$	$O^2(v_c)$	$E(v_c)$
v_c	$1 - \sigma\ (\sigma)$	$6\ (\sigma^2)$	$1\ (1/\sigma)$	$3\ (\sigma^2)$	$3\ (\sigma^3)$

For example, a good in the set $S(t_c)$ has for observation t_c a price equal to 1 and an endowment also equal to 1. For observation v_c the same good (which is in set $S(v_c)$) has price σ and endowment $1 - \sigma$. We assume that $\sigma > 0$ is a small but strict positive number.

Next, we also need to determine the incomes for each individual in a given observation. For each clause c we have exactly three variables x_i that are contained in a literal of c. The three individuals associated to these variables are the only individuals who have a strict positive income in the the observations t_c and v_c. That is, we set $I_{t_c}^i = p_{t_c} \varepsilon_{t_c}/3$ and $I_{v_c}^i = p_{v_c} \varepsilon_{v_c}/3$ for each i for which x_i is in a literal of clause c; we set $I_{t_c}^i = 0$ and $I_{v_c}^i = 0$ for all the other individuals. By construction we considered all observations and, as such, we determined all the individual incomes.

Finally, it is clear that the above construction of the sets of individuals, observations and goods, and the assignment of prices, endowments and incomes can be completed in a polynomial number of steps.

Step 2: M3SAT is satisfiable if the associated BMC-V problem is a "yes" instance

We will need the following two lemmas to prove the result.

Lemma 1. *For each pair of observations t_c and v_c (corresponding to a clause c), we have that if $\sigma > 0$ is small enough, then $p_{t_c}(\varepsilon_{t_c} - \varepsilon_{v_c}) > 0$.*

Proof. In order to show the required inequality, we construct Table 2 where each row or column corresponds with a class of goods for the observations v_c or t_c.

Table 2: Differences in expenditures for the different classes of goods

	$S(t_c)$	$S'(t_c)$	$O(t_c)$	$O^2(t_c)$	$E(t_c)$
$S(v_c)$	σ	/	/	/	/
$S'(v_c)$	/	0	/	/	/
$O(v_c)$	/	/	/	/	$\sigma^3(2)$
$O^2(v_c)$	/	/	/	/	0
$E(v_c)$	/	/	$\sigma^2(-8/3-\sigma)$	$\sigma^2(-8/3)$	0

The entries in the table give the difference in expenditures at price level p_{t_c} for the good common to both the row and column class. By construction, $S(t_c)$ only has goods in common with $S(v_c)$, and $S'(t_c)$ only with $S'(v_c)$. As such we only obtain a difference in expenditures in the two respective cells. For example the cell $\big(S(t_c), S(v_c)\big)$ only contains the good $g(t_c, v_c)$ and for this good we have that the expenditures in t_c minus the expenditures in v_c (at price level p_{t_c}) equals $1 \cdot (1 - 1 + \sigma) = \sigma$. A similar computation holds for cell $\big(S'(t_c), S'(v_c)\big)$.
Next, the group $O(v_c)$, again by construction, only has goods in common with $E(v_c)$. As above, one can then compute the difference in expenditures (at price level p_{t_c}): $\sigma^3(3-1)$.
Thirdly, the group $O^2(v_c)$ also only has goods in common with $E(v_c)$. To see this, assume on the contrary that $g(v_d, t_e) \in O^2(v_c) \cap O^2(t_c)$; the other case is dealt with by construction. This implies that $g(v_d, t_c) \in O(t_c)$ and $g(v_c, t_e) \in O(v_c)$. Now, assume w.l.o.g. that c has all its literals unnegated. This implies that both d and e have all there literals negated. However, then $g(v_d, t_e)$ is not defined in our list of goods, which is a contradiction. The difference in expenditures (at price level p_{t_c}) for the goods common to $O^2(v_c)$ and $E(v_c)$ is given by $\sigma^3(3-3)$. Finally, the group $E(v_c)$ has thus goods in common with $O(t_c), O^2(t_c)$ and $E(t_c)$, which leads to the respective differences in expenditures.
The total difference of expenditures, $p_{t_c}(\varepsilon_{t_c} - \varepsilon_{v_c})$, is of course given by summing for all goods the differences in Table 2. From this table we conclude that if σ is small enough, the

term in $\big(\mathsf{S}(t_c), \mathsf{S}(v_c)\big)$ dominates the expression $p_{t_c}(\varepsilon_{t_c} - \varepsilon_{v_c})$. This value is strict positive since $\sigma > 0$. $\qquad\square$

Lemma 2. *For each pair of observations v_c and t_d (corresponding to clauses c and d of opposite signature) we have that if $\sigma > 0$ is small enough, then $p_{v_c}(\varepsilon_{v_c}/3 - \varepsilon_{t_d}) > 0$.*

Proof. As similar reasoning as in Lemma 1 leads to Table 3. In this case, if σ approaches

Table 3: Differences in expenditures for the different classes of goods

	$\mathsf{S}(v_c)$	$\mathsf{S}'(v_c)$	$\mathsf{O}(v_c)$	$\mathsf{O}^2(v_c)$	$\mathsf{E}(v_c)$
$\mathsf{S}(t_d)$	/	$\sigma^2(1)$	/	/	/
$\mathsf{S}'(t_d)$	$\sigma(-17/3 - \sigma/3)$	$\sigma^2(-4)$	/	/	/
$\mathsf{O}(t_d)$	/	/	$(1/\sigma)(\sigma)$	$\sigma^2(2/3 + \sigma)$	/
$\mathsf{O}^2(t_d)$	/	/	0	$\sigma^2(2/3)$	$\sigma^3(2/3)$
$\mathsf{E}(t_d)$	/	/	/	$\sigma^2(-2)$	$\sigma^3(-2)$

zero, the dominating term in $p_{v_c}(\varepsilon_{v_c}/3 - \varepsilon_{t_d})$ is given by $\big(\mathsf{O}(v_c), \mathsf{O}(t_d)\big)$. Again this is strictly positive. $\qquad\square$

Now we are ready to show that M3SAT is satisfiable if the associated RATIONALIZABILITY problem is a "yes" instance. Consider an instance where the RATIONALIZABILITY answer is "yes". Set the value of the variables x_i such that for each literal in a clause c the value of this literal is equal to one if individual i directly revealed prefers t_c to v_c. Assign arbitrary values to all variables that are not determined in this way.

Lemma 1 guarantees that the RATIONALIZABILITY answer is "yes" only if for each clause c there is an individual i, with strict positive income, such that the individual directly revealed prefers t_c to v_c. Indeed, the available income for observation t_c is equally split over the 3 individuals linked to clause c; Lemma 1 then implies that for all decompositions of ε_{v_c}, we always have that at least one of these three individuals directly revealed prefers observation t_c over v_c. Hence each clause contains at least one literal with a value equal to one.

So the only thing remaining in order to show that M3SAT is satisfiable, is to verify that two opposing literals can not have the value one at the same time. Assume on the contrary that c contains a literal of variable x_i and d contains the opposite literal of the same variable x_i; that is, we assume that individual i directly reveals prefers t_c to v_c and t_d to v_d. A similar reasoning as above shows that Lemma 2 implies that this individual i also strict directly revealed prefers v_c over t_d and similarly v_d over t_c. As such we obtain for individual i the revealed preference cycle $(t_c, v_c), (v_c, t_d), (t_d, v_d), (v_d, t_c)$, which violates GARP since we have a cycle with strict direct revealed preferred relations. This gives us the desired contradiction and thus we conclude that M3SAT is satisfiable.

Step 3: if the M3SAT is satisfiable, the associated RATIONALIZABILITY problem is a "yes" instance

To see the reverse, we need to verify that for each satisfiable instance of M3SAT we can find a "yes" instance to the associated RATIONALIZABILITY problem. In order to do this, we need to specify, for a given an solution of M3SAT the consumption bundles q_t^i for each individual i and observation t in such a way that the conditions in Theorem 2 are satisfied.

Consider a satisfiable instance of M3SAT and for each clause, pick exactly one literal which has the value one; for an individual i and a clause c, we say $i \in I(c)$ if this literal corresponds to the variable x_i. We determine the consumption bundles according to whether i is in $I(c)$ or not. In the assignment given in Table 4 below, we only mention individuals with strict positive income, since the consumption for individuals with zero income must of course be zero for every good. Table 4 provides consumption bundles depending on the observation and the type of good. Further, for each clause c we pick one good out of $S'(v_c)$ and we call this good $g(v_c)$; this good will play a role in the proof of Fact 4. The prices are recalled between brackets.

Table 4: Consumption bundles for individuals with strict positive income

$i \in I(c)$						
	$S(t_c)$	$g(v_c)$	$S'(t_c)-\{g(v_c)\}$	$O(t_c)$	$O^2(t_c)$	$E(t_c)$
t_c	$1/3\ (1)$	$2\ (\sigma^2)$	$2\ (\sigma^2)$	$1/9-\sigma/3\ (\sigma^2)$	$1/9\ (\sigma^2)$	$1\ (\sigma^3)$
	$S(v_c)$	$g(v_c)$	$S'(v_c)-\{g(v_c)\}$	$O(v_c)$	$O^2(v_c)$	$E(v_c)$
v_c	$1/3-3\sigma\ (\sigma)$	$14/3\ (\sigma^2)$	$2\ (\sigma^2)$	$1/3\ (1/\sigma)$	$1\ (\sigma^2)$	$1\ (\sigma^3)$

$j \notin I(c)$						
	$S(t_c)$	$g(v_c)$	$S'(t_c)-\{g(v_c)\}$	$O(t_c)$	$O^2(t_c)$	$E(t_c)$
t_c	$1/3\ (1)$	$2\ (\sigma^2)$	$2\ (\sigma^2)$	$1/9-\sigma/3\ (\sigma^2)$	$1/9\ (\sigma^2)$	$1\ (\sigma^3)$
	$S(v_c)$	$g(v_c)$	$S'(v_c)-\{g(v_c)\}$	$O(v_c)$	$O^2(v_c)$	$E(v_c)$
v_c	$1/3+\sigma\ (\sigma)$	$2/3\ (\sigma^2)$	$2\ (\sigma^2)$	$1/3\ (1/\sigma)$	$1\ (\sigma^2)$	$1\ (\sigma^3)$

Once can easily verify that the sum of the above consumption bundles is indeed equal to the specified aggregate endowments in 1. Moreover the above specification is also compatible with our specification of the individual incomes (see Step 1). This shows that these bundles satisfy the first two conditions of Theorem 2. To verify the third condition, we need to establish the revealed preference relations for all individuals i.

Fact 1. *For all individuals $i \in \mathsf{I}(c)$, if $\sigma > 0$ is sufficiently small, then we have that i strict directly revealed prefers observation t_c to observation v_c.*

Proof. Given that $I^i_{t_c} = p_{t_c} \varepsilon_{t_c}/3$, we need to show that $p_{t_c}(\varepsilon_{t_c}/3 - q^i_{v_c}) > 0$. Table 5 is constructed as before.

Table 5: Difference in expenditures

	$\mathsf{S}(t_c)$	$\mathsf{S}'(t_c)$	$\mathsf{O}(t_c)$	$\mathsf{O}^2(t_c)$	$\mathsf{E}(t_c)$
$\mathsf{S}(v_c)$	3σ	/	/	/	/
$g(v_c)$	/	$\sigma^2(-8/3)$	/	/	/
$\mathsf{S}'(v_c) - \{g(v_c)\}$	/	0	/	/	/
$\mathsf{O}(v_c)$	/	/	/	/	$\sigma^3(2/3)$
$\mathsf{O}^2(v_c)$	/	/	/	/	0
$\mathsf{E}(v_c)$	/	/	$\sigma^2(-8/9 - \sigma/3)$	$\sigma^2(-8/9)$	0

For $\sigma > 0$ sufficiently small, the entry in $(\mathsf{S}(t_c), \mathsf{S}(v_c))$ dominates and this value is strictly positive. □

Fact 2. *Assume clauses c and d have opposite signature. For all individuals i with positive income in v_c, we have that if $\sigma > 0$ is sufficiently small, i strict directly revealed prefers observation v_c to observation t_d.*

Proof. Follows immediately from Lemma 2 since $I^i_{v_c} = p_{v_c} \varepsilon_{v_c}/3$. □

Fact 3. *For all individuals i with strict positive income in observation t_c (or v_c), then if $\sigma > 0$ is sufficiently small, we have that i strict directly revealed prefers observation t_c (or v_c) to any observation where i has zero income.*

Proof. Straightforward. □

Fact 4. *For all individuals i, if $\sigma > 0$ is sufficiently small, we have that i has no strict directly revealed preference comparisons besides those mentioned by Facts 1, 2 and 3.*

Proof. To prove this, we need to consider every possible comparison between two situations not mentioned in the previous facts. This is a long derivation which is similar to the ones we made before. Therefore, we leave it for the end of this Appendix. □

We can use these facts to show that every individual satisfies GARP, given that $\sigma > 0$ is sufficiently small. Suppose by contradiction that GARP is violated for individual i. This implies that i must have a revealed preference cycle containing at least one strict direct revealed strict preference comparison. Such a cycle cannot contain observations where i has zero income, since such observations cannot be revealed preferred to observations with strict positive

26

income. Further, it is impossible that the cycle only contains elements v_c for some clauses c, because Fact 4 implies that none of the observations v_c are revealed preferred to each other. As such, there must be a clause c such that the revealed preference cycle contains the element t_c. If the literal associated with x_i in c is zero,then $i \notin I(c)$ and Fact 4 implies t_c is not revealed preferred to any other observations where i has positive income. Hence, the literal associated with x_i in c must be equal to one.

Further, Fact 2 and 4 show that the only observations that can be strictly revealed preferred to t_c by i is an observation v_d where d and c have opposite signature. Also, as i has positive income in v_d, it must by construction be that d contains a literal associated with the variable x_i. Fact 4 implies that the only observation that can dominate v_d is the observation t_d and by Fact 1 this only holds if $i \in I(c)$, that is the literal associated with x_i equals one.

As such we obtain that both literals in clauses c and d associated with the variable x_i must be equal to one. This is a contradiction with the fact that c and d have opposite signature. Using the equivalence between GARP and the Afriat inequalities (Theorem 1), we therefore obtain that all the conditions of Theorem 2 are satisfied. Hence, we can conclude that RATIONALIZ-ABILITY is a "yes"instance.

Proof of Fact 4

Fact 3 implies that we only need to consider the pair of observations where in both observations we have a common individual with positive income. We have the following 13 restrictions that need to be satisfied to show that all strict direct revealed preference relations for individual i are captured by the first three facts. (We recall that the strict positive incomes are given by $I^i_{t_c} = p_{t_c} \varepsilon_{t_c} / 3$ and $I^i_{v_c} = p_{v_c} \varepsilon_{v_c} / 3$.)

- clause c

  (a) for $p_{v_c}(\varepsilon_{v_c}/3 - q^i_{t_c}) < 0.$
  (b) for $j \notin I(c): p_{t_c}(\varepsilon_{t_c}/3 - q^j_{v_c}) < 0 .$

- clause c and d have opposite signature

  (c) for $i \in I(c)$ $p_{t_d}(\varepsilon_{t_d}/3 - q^i_{v_c}) < 0.$
  (d) for $j \notin I(c): p_{t_d}(\varepsilon_{t_d}/3 - q^j_{v_c}) < 0.$
  (e) for $p_{t_c}(\varepsilon_{t_c}/3 - q^i_{t_d}) < 0.$
  (f) for $i \in I(d): p_{v_c}(\varepsilon_{v_c}/3 - q^i_{v_d}) < 0.$
  (g) for $j \notin I(d): p_{v_c}(\varepsilon_{v_c}/3 - q^j_{v_d}) < 0.$

- clause c and d have the same signature

  (h) for $p_{v_c}(\varepsilon_{v_c}/3 - q^i_{t_d}) < 0.$

(i) for $i \in I(d)$: $p_{v_c}(\varepsilon_{v_c}/3 - q^i_{v_d}) < 0$.

(j) for $j \notin I(d)$: $p_{v_c}(\varepsilon_{v_c}/3 - q^j_{v_d}) < 0$.

(k) for $i \in I(c)$: $p_{t_d}(\varepsilon_{t_d}/3 - q^i_{v_c}) < 0$.

(l) for $j \notin I(c)$: $p_{t_d}(\varepsilon_{t_d}/3 - q^j_{v_c}) < 0$.

(m) for $p_{t_c}(\varepsilon_{t_c}/3 - q^i_{t_d}) < 0$.

The following thirteen tables of differences of expenditures show that the above inequalities hold if σ is sufficiently small.

case a.

$$p_{v_c}(\varepsilon_{v_c}/3 - \varepsilon_{t_c}/3) < 0$$

	$S(v_c)$	$S'(v_c)$	$O(v_c)$	$O^2(v_c)$	$E(v_c)$
$S(t_c)$	$\sigma(-\sigma/3)$	/	/	/	/
$S'(t_c)$	/	0	/	/	/
$O(t_c)$	/	/	/	/	$\sigma^3(8/9+\sigma/3)$
$O^2(t_c)$	/	/	/	/	$\sigma^3(8/9)$
$E(t_c)$	/	/	$1/\sigma(-2/3)$	0	0

For small σ, the term in $(O(v_c), E(t_c))$ dominates. If this cell is empty, we are led to the value in cell $(S(v_c), S(t_c))$.

case b.

$$j \notin I(c): p_{t_c}(\varepsilon_{t_c}/3 - q^j_{v_c}) < 0$$

	$S(t_c)$	$S'(t_c)$	$O(t_c)$	$O^2(t_c)$	$E(t_c)$
$S(v_c)$	$-\sigma$	/	/	/	/
$g(v_c)$	/	$\sigma^2(4/3)$	/	/	/
$S'(v_c) - \{g(v_c)\}$	/	0	/	/	/
$O(v_c)$	/	/	/	/	$\sigma^3(2/3)$
$O^2(v_c)$	/	/	/	/	0
$E(v_c)$	/	/	$\sigma^2(-8/9-\sigma/3)$	$\sigma^2(-8/9)$	0

In this case, the term in cell $(S(t_c), S(v_c))$ dominates.

case c.

$$i \in I(c): p_{t_d}(\varepsilon_{t_d}/3 - q^i_{v_c}) < 0.$$

	$S(t_d)$	$S'(t_d)$	$O(t_d)$	$O^2(t_d)$	$E(t_d)$
$S(v_c)$	/	$\sigma^2(5/3+3\sigma)$	/	/	/
$g(i,v_c)$	$-13/3$	$\sigma^2(-8/3)$	/	/	/
$S'(v_c) - \{g(i,v_c)\}$	$-5/3$	0	/	/	/
$O(v_c)$	/	/	$\sigma^2(-2/9-\sigma/3)$	$\sigma^2(-1/9)$	/
$O^2(v_c)$	/	/	$\sigma^2(-8/9-\sigma/3)$	$\sigma^2(-8/9)$	/
$E(v_c)$	/	/	/	/	0

If $S(t_d) = g(v_c)$, we need to put the entries in the cells $(S(t_d), S'(v_c) - \{g(v_c)\})$ and $(S'(t_d), g(v_c))$ equal to $/$. In that case the cell $(S(t_d), g(v_c))$ dominates. If $S(t_d) \neq g(v_c)$, we need to put the entry in the cell $(S(t_d), g(v_c))$ equal to $/$ and then the cell $(S(t_d), S'(v_c) - g(v_c))$ dominates.

case d.

$$j \notin I(c): p_{t_d}(\varepsilon_{t_d}/3 - q_{v_c}^j) < 0.$$

	$S(t_d)$	$S'(t_d)$	$O(t_d)$	$O^2(t_d)$	$E(t_d)$
$S(v_c)$	$/$	$\sigma^2(5/3 - \sigma)$	$/$	$/$	$/$
$g(v_c)$	$-1/3$	$\sigma^2(4/3)$	$/$	$/$	$/$
$S'(v_c) - \{g(v_c)\}$	$-5/3$	o	$/$	$/$	$/$
$O(v_c)$	$/$	$/$	$\sigma^2(-2/9 - \sigma/3)$	$\sigma^2(-2/9)$	$/$
$O^2(v_c)$	$/$	$/$	$\sigma^2(-8/9 - \sigma/3)$	$\sigma^2(-8/9)$	$/$
$E(v_c)$	$/$	$/$	$/$	$/$	o

The same applies as in case c: If $S(t_d) = g(v_c)$, we need to put the entries in the cells $(S(t_d), S'(v_c) - \{g(v_c)\})$ and $(S'(t_d), g(v_c))$ equal to $/$. In that case the cell $(S(t_d), g(v_c))$ dominates. If $S(t_d) \neq g(v_c)$, we need to put the entry in the cell $(S(t_d), g(v_c))$ equal to $/$ and then the cell $(S(t_d), S'(v_c) - g(v_c))$ dominates.

case e.

$$p_{t_c}(\varepsilon_{t_c}/3 - \varepsilon_{t_d}/3) < 0.$$

	$S(t_c)$	$S'(t_c)$	$O(t_c)$	$O^2(t_c)$	$E(t_c)$
$S(t_d)$	$/$	$\sigma^2(5/3)$	$/$	$/$	$/$
$S'(t_d)$	$-5/3$	o	$/$	$/$	$/$
$O(t_d)$	$/$	$/$	$/$	$/$	$\sigma^3(8/9 + \sigma/3)$
$O^2(t_d)$	$/$	$/$	$/$	$/$	$\sigma^3(8/9)$
$E(t_d)$	$/$	$/$	$\sigma^2(-8/9 - \sigma/3)$	$\sigma^2(-8/9)$	o

For this case, the term in the cell $(S(t_c), S'(t_d))$ dominates.

case f.

$$i \in I(d): p_{v_c}(\varepsilon_{v_c}/3 - q_{v_d}^i) < 0.$$

	$S(v_c)$	$S'(v_c)$	$O(v_c)$	$O^2(v_c)$	$E(v_c)$
$S(v_d)$	$/$	$\sigma^2(5/3 + 3\sigma)$	$/$	$/$	$/$
$g(v_d)$	$\sigma(-13/3 - \sigma/3)$	$\sigma^2(-8/3)$	$/$	$/$	$/$
$S'(v_d) - \{g(v_d)\}$	$\sigma(-5/3 - \sigma/3)$	o	$/$	$/$	$/$
$O(v_d)$	$/$	$/$	$/$	$/$	$\sigma^3(2/3)$
$O^2(v_d)$	$/$	$/$	$/$	$/$	o
$E(v_d)$	$/$	$/$	$1/\sigma(-2/3)$	o	o

It is the entry in the cell $(O(v_c), E(v_d))$ that dominates.

case g.

$$j \notin I(d): p_{v_c}(\varepsilon_{v_c}/3 - q_{v_d}^j) < 0.$$

	$S(v_c)$	$S'(v_c)-\{g(v_d)\}$	$O(v_c)$	$O^2(v_c)$	$E(v_c)$
$S(v_d)$	/	$\sigma^2(5/3-\sigma)$	/	/	/
$g(v_d)$	$\sigma(1/3-\sigma/3)$	$\sigma^2(4/3)$	/	/	/
$S'(v_d)-\{g(v_d)\}$	$\sigma(-5/3-\sigma/3)$	0	/	/	/
$O(v_d)$	/	/	/	/	$\sigma^3(2/3)$
$O^2(v_d)$	/	/	/	/	0
$E(v_d)$	/	/	$1/\sigma(-2/3)$	0	0

Again, a distinction has to be made whether $g(v_d)=S(v_c)$ or not. However, in any case it is the entry in the cell $\big(O(v_c),E(v_d)\big)$ that dominates.

case h.

$$p_{v_c}(\varepsilon_{v_c}/3-\varepsilon_{t_d}/3)<0.$$

	$S(v_c)$	$S'(v_c)$	$O(v_c)$	$O^2(v_c)$	$E(v_c)$
$S(t_d)$	/	$\sigma^2(5/3)$	/	/	/
$S'(t_d)$	$\sigma(-5/3-\sigma/3)$	0	/	/	/
$O(t_d)$	/	/	/	/	$\sigma^3(8/9+\sigma/3)$
$O^2(t_d)$	/	/	/	/	$\sigma^3(8/9)$
$E(t_d)$	/	/	$1/\sigma(-2/3)$	0	0

In this case, entry $\big(O(v_c),E(t_d)\big)$ dominates.

case i.

$$i\in I(d):\; p_{v_c}(\varepsilon_{v_c}/3-q^i_{v_d})<0.$$

	$S(v_c)$	$S'(v_c)$	$O(v_c)$	$O^2(v_c)$	$E(v_c)$
$S(v_d)$	/	$\sigma^2(5/3+3\sigma)$	/	/	/
$g(v_d)$	$\sigma(-13/3-\sigma/3)$	$\sigma^2(-8/3)$	/	/	/
$S'(v_d)-\{g(v_d)\}$	$\sigma(-5/3-\sigma/3)$	0	/	/	/
$O(v_d)$	/	/	/	$\sigma^2(-2/3)$	/
$O^2(v_d)$	/	/	$1/\sigma(-2/3)$	0	/
$E(v_d)$	/	/	/	/	0

We can make a distinction between the cases $g(v_d)\in S(v_c)$ and $g(v_d)\notin S(v_c)$. Anyhow, the term that dominate is $\big(O(v_c),O^2(v_d)\big)$.

case j.

$$j\notin I(d):\; p_{v_c}(\varepsilon_{v_c}/3-q^j_{v_d})<0.$$

	$S(v_c)$	$S'(v_c)$	$O(v_c)$	$O^2(v_c)$	$E(v_c)$
$S(v_d)$	/	$\sigma^2(5/3 - \sigma)$	/	/	/
$g(v_d)$	$\sigma(-1/3 - \sigma/3)$	$\sigma^2(4/3)$	/	/	/
$S'(v_d) - \{g(v_d)\}$	$\sigma(-5/3 - \sigma/3)$	0	/	/	/
$O(v_d)$	/	/	/	$\sigma^2(-2/3)$	/
$O^2(v_d)$	/	/	$1/\sigma(-2/3)$	0	/
$E(v_d)$	/	/	/	/	0

Again, we can make a distinction between the cases $g(v_d) \in S(v_c)$ and $g(v_d) \notin S(v_c)$. Anyhow, the term that dominate is $\big(O(v_c), O^2(v_d)\big)$.
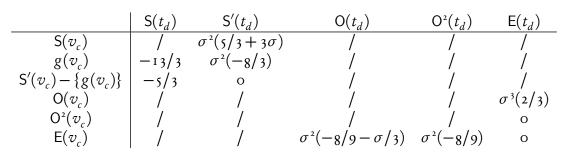
case k.

$$i \in I(c): p_{t_d}(\varepsilon_{t_d}/3 - q_{v_c}^i) < 0.$$

	$S(t_d)$	$S'(t_d)$	$O(t_d)$	$O^2(t_d)$	$E(t_d)$
$S(v_c)$	/	$\sigma^2(5/3 + 3\sigma)$	/	/	/
$g(v_c)$	$-13/3$	$\sigma^2(-8/3)$	/	/	/
$S'(v_c) - \{g(v_c)\}$	$-5/3$	0	/	/	/
$O(v_c)$	/	/	/	/	$\sigma^3(2/3)$
$O^2(v_c)$	/	/	/	/	0
$E(v_c)$	/	/	$\sigma^2(-8/9 - \sigma/3)$	$\sigma^2(-8/9)$	0

If $g(v_c) \in S(t_d)$, then the entries in $\big(S(t_d), S'(v_c) - \{g(v_c)\}\big)$ and $\big(S'(t_d), g(v_c)\big)$ are empty and the leading term is given by $\big(S(t_d), g(v_c)\big)$. if $g(v_c) \notin S(t_d)$, then the cell $\big(S(t_d), g(v_c)\big)$ is empty and the leading term is $\big(S(t_d), S'(v_c) - \{g(v_c)\}\big)$.

case l.

$$j \notin I(c): p_{t_d}(\varepsilon_{t_d}/3 - q_{v_c}^j) < 0.$$

	$S(t_d)$	$S'(t_d)$	$O(t_d)$	$O^2(t_d)$	$E(t_d)$
$S(v_c)$	/	$\sigma^2(5/3 - \sigma)$	/	/	/
$g(v_c)$	$-1/3$	$\sigma^2(4/3)$	/	/	/
$S'(v_c) - \{g(v_c)\}$	$-5/3$	0	/	/	/
$O(v_c)$	/	/	/	/	$\sigma^3(2/3)$
$O^2(v_c)$	/	/	/	/	0
$E(v_c)$	/	/	$\sigma^2(-8/9 - \sigma/3)$	$\sigma^2(-8/9)$	0

If $g(v_c) \in S(t_d)$, then the entries in $\big(S(t_d), S'(v_c) - \{g(v_c)\}\big)$ and $\big(S'(t_d), g(v_c)\big)$ are empty and the leading term is given by $\big(S(t_d), g(v_c)\big)$. if $g(v_c) \notin S(t_d)$, then the cell $\big(S(t_d), g(v_c)\big)$ is empty and the leading term is $\big(S(t_d), S'(v_c) - \{g(v_c)\}\big)$.

case m.

$$p_{t_c}(\varepsilon_{t_c}/3 - \varepsilon_{t_d}/3) < 0.$$

	$\mathsf{S}(t_c)$	$\mathsf{S}'(t_c)$	$\mathsf{O}(t_c)$	$\mathsf{O}^2(t_c)$	$\mathsf{E}(t_c)$
$\mathsf{S}(t_d)$	/	$\sigma^2(5/3)$	/	/	/
$\mathsf{S}'(t_d)$	$-5/3$	o	/	/	/
$\mathsf{O}(t_d)$	/	/	/	$\sigma^2(\sigma/3)$	/
$\mathsf{O}^2(t_d)$	/	/	$\sigma^2(-\sigma/3)$	o	/
$\mathsf{E}(t_d)$	/	/	/	/	o

The term that dominates in this table is $\big(\mathsf{S}(t_c),\mathsf{S}'(t_d)\big)$.

References

Afriat, S. N., 1967. The construction of utility functions from expenditure data. International Economic Review 8, 67–77.

Andreoni, J., Harbaugh, W., 2006. Power indices for revealed preference tests. University of Winconsin-Madison Department od Economics Working Paper 2005-10.

Bachmann, R., 2006. Testable implications of Pareto efficiency and individual rationality. Economic Theory 29, 489–504.

Bergstrom, T., 1976. Collective choice and the Lindahl allocation method. University of California at Santa Barbara, Economics Working Paper Series 1976A.

Bronars, S., 1987. The power of nonparametric tests of preference maximization. Econometrica 55, 693–698.

Brown, D. J., Kannan, R., 2008. Two algorithms for solving the Walrasian equilibrium inequalities, Chapter 6 in *Computational Aspects of General Equilibrium Theory*. Springer-Verlag, New York and Berlin.

Brown, D. J., Lee, Y.-H. A., 2008. Competition, Consumer Welfare, and the Social Cost of Monopoly, Chapter 5 in *Computational Aspects of General Equilibrium Theory*. Springer-Verlag, New York and Berlin.

Brown, D. J., Matzkin, R. L., 1996. Testable restrictions on the equilibrium manifold. Econometrica, 1249–1262.

Carvajal, A., 2003. Testable restrictions on the equilibrium manifold under random utility. Borradores de Economía 233.

Carvajal, A., 2005. Testable restrictions of general equilibrium in production economies. Royal Holloway University of London Discussion paper 2005-01.

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

Cherchye, L., De Rock, B., Sabbe, J., Vermeulen, F., 2008. Nonparametric tests of collective rational consumption bhevior: An integer programming procedure. Journal of Econometrics 147, 258–265.

Cherchye, L., De Rock, B., Vermeulen, F., 2007. The revealed preference approach to collective consumption behavior: testing, recovery and welfare analysis. CES Discussion paper 07.24.

Cherchye, L., De Rock, B., Vermeulen, F., 2007a. The collective model of household consumption: a nonparametric characterization. Econometrica 75, 553–574.

Chiappori, P. A., 1997. Introducing household production in collective models of labor supply. Journal of Political Economy 100, 437–467.

Chiappori, P. A., Ekeland, I., Kübler, F., Polemarchakis, H., 2004. Testable implications of general equilibrium theory: a differentiable approach. Journal of Mathematical Economics 40, 105–119.

Chiapppori, P. A., Ekeland, I., Kübler, F., Polemarchakis, H. M., 1999. The identification of preferences from equilibrium prices. CORE Discussion Paper no 2000/24.

Deb, R., 2008. An efficient nonparametric test of the collective household model. Yale University working paper.

Deb, R., 2009. A testable model of consumption with externalities. Journal of Economic Theory forthcomming.

Diewert, W. E., 1973. Afriat and revealed preference theory. The Review of Economic Studies 40, 419–425.

Garey, M. R., Johnson, D. S., 1979. Computers and Intractability. Freeman, New York.

Kübler, F., 2003. Observable restrictions of general equilibrium models with financial markets. Journal of Economic Theory 110, 137–153.

Kübler, F., Schmedders, K., 2008. Approximate Versus Exact Equilibria in Dynamic Economies, Chapter 10 in *Computational Aspects of General Equilibrium Theory*. Springer-Verlag, New York and Berlin.

Rizvi, A. T., 2006. The Sonnenschein-Mantel-Debreu result after thirty years. History of Political Economy Supplement 38, 228–245.

Sethuraman, J., Piaw, T., Vohra, R., 2003. Integer programming and Arrovian social welfare functions. Mathematics of Operations Research 28, 309–326.

Snyder, S. K., 1999. Testable restrictions of Pareto optimal public good provision. Journal of Public Economics 71, 97–119.

Snyder, S. K., 2004. Observable implications of equilibrium behavior on finite data. Journal of Mathematical Economics 40, 165–176.

Talla Nobibon, Cherchye, L., De Rock, B., Sabbe, J., Spiekma, F., 2008. Heuristics for deciding collectively rational consumption behavior. CES Discussion paper 08.24.

Varian, H., 1982. The nonparametric approach to demand analysis. Econometrica 50, 945–972.

Varian, H., 1984. The nonparametric approach to production analysis. Econometrica 52, 579–697.

Varian, H., 1985. Nonparametric analysis of optimizing behavior with measurement error. Journal of Econometrics 30, 445–458.

Varian, H., 1990. Goodness of fit in optimizing models. Journal of Econometrics 46, 125–140.

Varian, H., 2006. Samuelsonian Economics and the 21st Century. Oxford University press, Ch. Revealed preferences.

Warshall, S., 1962. A theorem of boolean matrices. Journal of the Americal Association of Computing Machinery 9, 11–12.