Journal of Mathematical Economics 58 (2015) 82–91

Contents lists available at ScienceDirect

Journal of Mathematical Economics

journal homepage: www.elsevier.com/locate/jmateco

Complexity results for the weak axiom of revealed preference for collective consumption models

Bart Smeulders[a,*], Laurens Cherchye[b], Bram De Rock[c], Frits C.R. Spieksma[a], Fabrice Talla Nobibon[d]

[a] ORSTAT, KU Leuven, Belgium
[b] Center for Economic Studies, KU Leuven, Belgium
[c] ECARES, Université Libre de Bruxelles, Belgium
[d] Fedex Express Europe, Middle East, Indian Subcontinent & Africa, Belgium

ARTICLE INFO

Article history:
Received 28 October 2014
Accepted 31 March 2015
Available online 11 April 2015

Keywords:
Revealed preferences
Collective model
Computational complexity

ABSTRACT

The purpose of this paper is to establish the complexity of alternative versions of the weak axiom of revealed preference (WARP) for collective consumption models. In contrast to the unitary consumption model, these collective models explicitly take the multi-member nature of the household into account. We consider the three collective settings that are most often considered in the literature. We start with the private setting in which all goods are privately consumed by the household members. Next, we consider the public setting in which all goods are publicly consumed inside the household. Finally, we also consider the general setting where no information on the (private or public) nature of goods consumed in the household is available. We prove that the collective version of WARP is NP-hard to test for both the private and public settings. Surprisingly, we also find for the general setting that the collective version of WARP is easy to test for two-member households.

1. Introduction

Modeling and analyzing household consumption behavior is a fundamental research topic in microeconomics. For a long time, the standard model in empirical consumption analysis was the so-called *unitary* model, which treats the household as a single decision making unit. However, by now it is well established that Chiappori (1988)'s *collective* model of household consumption is both conceptually and empirically a more attractive alternative for analyzing the consumption behavior of multi-member households (see, for example, Vermeulen, 2002 for an overview of the relevant literature). This collective model assumes that the different household members are endowed with individual preferences defined over privately and publicly consumed goods (inside the household). These members then enter into a decision process of which the outcome is assumed to obtain a Pareto optimal allocation (of the aggregate household budget).

In the tradition of Afriat (1967) and Varian (1982), we are interested in the revealed preference characterization of collective models. Such a revealed preference characterization does not rely on any functional specification regarding the household consumption process, and starts directly from the observed finite set of prices and quantities. Varian (1982) introduced the revealed preference axioms that summarize the empirical implications of theoretical consumption models for single-member households. Basically, consistency with the unitary model requires the observed consumption data to obey the *strong axiom of revealed preference* (SARP). More recently, Cherchye et al. (2011) provided a revealed preference characterization of collective models for multi-member households, which implies a multi-member version of SARP.[1]

To date, this collective extension of SARP has received extensive analysis. Most notably, it has been shown that testing the SARP conditions for collective models is NP-complete, even for households with only two members (Deb, 2010; Talla Nobibon et al.,

* Correspondence to: Naamsestraat 69, 3000 Leuven, Belgium. Tel.: +32 1632692.
E-mail address: bart.smeulders@kuleuven.be (B. Smeulders).

http://dx.doi.org/10.1016/j.jmateco.2015.03.010

[1] See also Peters and Wakker (1994), Varian (2006), Cherchye et al. (2007), Cherchye et al. (2010) and Cherchye et al. (2013) for more discussion. To be precise, Cherchye et al. (2011) actually characterized the collective model in terms of the *generalized axiom of revealed preference* (GARP) rather than SARP. But their results are easily translated towards *SARP*. See, for example, Varian (1982) for a detailed discussion on the subtle difference between GARP and SARP.

2012; Talla Nobibon and Spieksma, 2010). Importantly, this contrasts with complexity results for the unitary model, for which the SARP conditions can be tested in polynomial time (Piaw and Vohra, 2003; Talla Nobibon et al., 2014; Varian, 1982).

In this paper, we complement these existing results by considering the collective version of the so-called *weak axiom of revealed preference* (WARP). Basically, WARP coincides with SARP except from the fact that it does not require revealed preferences to be transitive. Thus, in general, WARP and SARP are not equivalent to each other. However, it turns out that in practical applications they often have identical empirical implications, i.e. most data that satisfy WARP also satisfy SARP. Putting it differently, in empirical work transitivity usually plays little role when testing data consistency with revealed preference axioms. This last observation is an important one in view of practical tests of the collective model (SARP) restrictions, as transitivity restrictions usually occupy a lot of computation time.

This also directly motivates the purpose of the current paper, which focuses on the computational complexity of the collective WARP conditions. Tools of computational complexity are increasingly used to investigate computational properties of various problems in economic behavior; we restrict ourselves here to mentioning work in goodness-of-fit models (Echenique et al., 2011; Smeulders et al., 2014) and rationalizing choice behavior (Apesteguia and Ballester, 2010; Demuynck, 2011). Essentially, we will evaluate whether the computational hardness of the collective revealed preference conditions can be mitigated by dropping the transitivity requirement. In particular, our following analysis will consider the WARP characterization of three collective consumption settings: (i) the *private* setting where all goods are consumed privately without externalities, (ii) the *public* setting, where all goods are publicly consumed inside the household, and (iii) a *general* setting where no information on the (private or public nature) of the goods is available.

Our main findings can be summarized as follows. A first "negative" conclusion will be that testing the collective WARP conditions is computationally difficult (i.e. NP-complete) for the private and public settings. In these cases, dropping transitivity does not solve the hardness problem associated with the collective SARP conditions. However, as a second "positive" conclusion, we also show that testing collective WARP for two members is computationally easy for the general setting. Here, we can effectively test consistency with the collective consumption model in an efficient way (i.e. in polynomial time) if we omit transitivity. (As we will indicate, for this general setting the complexity in the case of three or more members remains an open question.)

The remainder of the paper unfolds as follows. Section 2 presents our basic set-up. Sections 3–5 contain our main complexity results (for, respectively, the private, public and general settings). Section 6 concludes.

2. Set-up

To set the stage, we will start by fixing our basic notation. Next, we will introduce the WARP conditions that apply to single-member households (which are essentially the conditions that apply to the unitary consumption model). The following sections will be more specific on the collective models (and corresponding WARP conditions) that form the central focus of our analysis.

2.1. Notation

We consider multi-member households that take consumption decisions over m commodities (or goods). These goods can be consumed either privately (with or without externalities) or publicly. More precisely, *private* consumption of a good means that the consumption by one household member affects the supply available for the other household members (e.g. drinking water can only be consumed privately). Next, consumption *externalities* refer to the fact that one household member gets utility from another household member's private consumption (e.g. a wife enjoys her husbands nice clothes). Finally, *public* consumption of a good means that consumption of that good by one household member does not affect the supply available for the other household members, and no one can be excluded from consuming the good (e.g. the rent of a shared house represents public consumption).

The collective models of household consumption explicitly recognize the individual preferences of the household members. These preferences may depend on the private quantities (with or without externalities), the public quantities, or both. Throughout, we assume that preferences of the household members can be represented by a well-behaved (i.e. continuous, positive monotonic and concave) utility function. The following sections will define explicit specifications of these member-specific utility functions for alternative collective consumption models.

We assume a setting in which the empirical analyst observes n household decisions resulting in consumption quantity bundles $q_t := (q_{t,1}, \dots, q_{t,m}) \in \mathbb{R}^m_+$, with corresponding prices $p_t := (p_{t,1}, \dots, p_{t,m}) \in \mathbb{R}^m_{++}$, $t = 1, \dots, n$. The component $q_{t,i}$ (respectively $p_{t,i}$), for $i = 1, \dots, m$, corresponds to the quantity of good i bought by the household (respectively, the unit price of good i) at the time of observation t. Note that the scalar product $p'q$ represents the total outlay for bundle $q \in \mathbb{R}^m_+$ at the prices $p \in \mathbb{R}^m_{++}$. For ease of notation, we will write this scalar product simply as pq. We denote the set of observations by $S := \{(p_t, q_t) : t \in N\}$, where $N := \{1, \dots, n\}$, and we refer to S as the *dataset*. For ease of exposition, throughout this paper, we use $t \in N$ to refer to the observation (p_t, q_t).

2.2. WARP *for single-member households*

Samuelson (1938) originally introduced the WARP condition for single-member households. It defines a necessary requirement for the existence of a single well-behaved utility function that is consistent with the observed dataset $S := \{(p_t, q_t) : t \in N\}$. More precisely, there can only exist a well-behaved utility function that is maximized by each observed bundle q_t subject to the corresponding budget constraint (defined by the prices p_t and the budget $p_t q_t$), if the dataset S is consistent with WARP. As indicated in the Introduction, the necessary WARP requirement differs from the necessary and sufficient SARP requirement in that it does not require (revealed) preferences to be transitive.

The formal definition of WARP is as follows.

Definition 1. Let $S := \{(p_t, q_t) : t \in N\}$ be an observed dataset.

1. Bundle q_s is *directly revealed preferred* over bundle q_t if and only if $p_s q_s \geq p_s q_t$.
2. S satisfies WARP if and only if, for all observations $s, t \in N$, when $q_s \neq q_t$ and q_s is directly preferred over q_t, then $p_t q_t < p_t q_s$.

In words, the bundle q_s is directly revealed preferred over the bundle q_t if the bundle q_s was chosen, while the bundle q_t was also affordable (at the prices p_s). Next, if a bundle q_s is directly revealed preferred over q_t, then it cannot be that q_t is also directly revealed preferred over q_s (unless both bundles are identical).

In the following sections, we will extend this single-member WARP requirement to alternative collective settings. Collective consumption models pertain to multi-member households, in which each member has his or her own well-behaved utility function. Therefore, our consistency conditions for collective models will basically impose WARP requirements for the (multiple) individual household members. The specific form of these WARP requirements will depend on the collective model at hand.

3. The collective consumption model with only private consumption and no externalities

In the first collective consumption model that we study, we assume that all goods are consumed privately without externalities. In other words, the member-specific utility functions only depend on the private goods consumed by that member. To facilitate our discussion, we will mainly focus on two-member households in what follows. However, as we will also indicate in Theorem 2, our NP-completeness result for two-member households can easily be generalized to households with k members ($k \geq 2$).

Because a typical dataset only contains information on consumption quantities that apply to the aggregate household level, we have to deal with the fact that we do not know which fraction of the observed bundle q_t is consumed by each individual household member. To this end, we consider, for each observation $t \in N$, a *feasible personalized quantity vector* (q_t^1, q_t^2), which describes the division of the goods over the two household members. Since the true split up of q_t is unobserved, we clearly need to consider all possible feasible personalized quantity vectors.[2] For each member ℓ ($\ell = 1, 2$) we define the *personalized consumption dataset* by $S_\ell = \{(p_t, q_t^\ell) : t \in N\}$.

The extension of WARP to this collective consumption model is then as follows.

Definition 2 (*Private 2*-WARP)**.** Let $S = \{(p_t, q_t) : t \in N\}$ be a dataset of a two-member household. We say that S is consistent with *private* 2-WARP if and only if:

(i) For each $t \in N$ there exist $q_t^1, q_t^2 \in \mathbb{R}_+^m$ such that $q_t = q_t^1 + q_t^2$, and
(ii) For each member $\ell \in \{1, 2\}$, the set $S_\ell = \{(p_t, q_t^\ell) : t \in N\}$ satisfies WARP.

This problem can be rephrased as the following decision problem:

Problem: private 2-WARP.

Instance: A dataset $S = \{(p_t, q_t) : t \in N\}$.

Question: Do there exist $q_t^1, q_t^2 \in \mathbb{R}_+^m$ satisfying $q_t = q_t^1 + q_t^2$ for each $t \in N$ such that for $\ell = 1, 2$, the set $S_\ell = \{(p_t, q_t^\ell) : t \in N\}$ satisfies WARP?

It turns out that answering this question is NP-complete even for the case under consideration, with only two members in the household.

Theorem 1. *Testing private* 2-WARP *is NP-complete.*

Proof. We use a reduction from MONOTONE NOT-ALL-EQUAL 3-SAT, which is known to be NP-complete (Garey and Johnson, 1979).

Instance: A set of variables $X = \{x_1, x_2, \ldots, x_n\}$ and a set of clauses $C = \{c_1, c_2, \ldots, c_m\}$ with each clause consisting of 3 non-negated literals.

Question: Does there exist a truth-assignment so that for each clause, either one or two of the literals are TRUE?

It is not difficult to see that private 2-WARP belongs to the class NP. The rest of this proof is structured as follows: given an arbitrary instance of MNAE 3-SAT, we first build an instance of private 2-WARP and next, we prove that we have a YES instance of MNAE 3-SAT if and only if the constructed instance of private 2-WARP is a YES instance.

Consider an arbitrary instance $X = \{x_1, x_2, \ldots, x_n\}$ and $C = \{c_1, c_2, \ldots, c_m\}$ of MNAE 3-SAT. We build an instance of private 2-WARP using $3n + 4$goods and $2n + 2m + 3$ observations. We next describe the quantity and the price of goods for each observation. We use $\epsilon = \frac{1}{4n}$ and $M = n + 1$. The first block of $2n$ observations corresponds to the variables and is given by:

$$q_1 = (0, 0, |0, 0, |0, \ldots, 0, |1, 0, \ldots, 0, |0, 0, \ldots, 0);$$
$$p_1 = (M, M, |M, M, |1, \epsilon, \ldots, \epsilon, |1, M, \ldots, M, |\epsilon, M, \ldots, M)$$
$$q_2 = (0, 0, |0, 0, |0, \ldots, 0, |0, 1, \ldots, 0, |0, 0, \ldots, 0);$$
$$\vdots$$
$$q_n = (0, 0, |0, 0, |0, \ldots, 0, |0, 0, \ldots, 1, |0, 0, \ldots, 0);$$
$$p_2 = (M, M, |M, M, |\epsilon, 1, \ldots, \epsilon, |M, 1, \ldots, M, |M, \epsilon, \ldots, M)$$
$$\vdots$$
$$p_n = (M, M, |M, M, |\epsilon, \epsilon, \ldots, 1, |M, M, \ldots, 1, |M, M, \ldots, \epsilon)$$
$$q_{n+1} = (0, 0, |0, 0, |0, \ldots, 0, |0, 0, \ldots, 0, |1, 0, \ldots, 0);$$
$$\vdots$$
$$q_{2n} = (0, 0, |0, 0, |0, \ldots, 0, |0, 0, \ldots, 0, |0, \ldots, 0, 1);$$
$$p_{n+1} = (M, M, |M, M, |M, \ldots, M, |\epsilon, M, \ldots, M, |1, M, \ldots, M)$$
$$\vdots$$
$$p_{2n} = (M, M, |M, M, |M, \ldots, M, |M, M, \ldots, \epsilon, |M, M, \ldots, 1)$$

The second block of $2m$ observations corresponds to the clauses. For each clause $c_a = \{x_i, x_j, x_k\}$, we have the observations $2n + a$ and $2n + m + a$ ($a = 1, \ldots, m$).

$$q_{2n+a} = (1, 0, |0, 0, |0, \ldots, 0, |0, \ldots, 0, |0, \ldots, 0);$$
$$p_{2n+a} = (1, \epsilon, |M, M, | \{\epsilon, 1\} |M, \ldots, M, |M, \ldots, M)$$
$$q_{2n+m+a} = (0, 1, |0, 0, |0, \ldots, 0, |0, \ldots, 0, |0, \ldots, 0);$$
$$p_{2n+m+a} = (\epsilon, 1, |M, M, | \{\epsilon, 1\} |M, \ldots, M, |M, \ldots, M).$$

The prices of the goods corresponding to variables x_i, x_j and x_k equal 1, and the prices of the goods corresponding to other variables equal ϵ. Finally, we have observations $2n + 2m + 1$, $2n + 2m + 2$, $2n + 2m + 3$.

$$q_{2n+2m+1} = (0, 0, |1, 0, |0, \ldots, 0, |0, \ldots, 0, |0, \ldots, 0);$$
$$p_{2n+2m+1} = (M, M, |1, \epsilon, |\epsilon, \ldots, \epsilon, |M, \ldots, M, |M, \ldots, M)$$
$$q_{2n+2m+2} = (0, 0, |0, 1, |0, \ldots, 0, |0, \ldots, 0, |0, \ldots, 0);$$
$$p_{2n+2m+2} = (M, M, |\epsilon, 1, |\epsilon, \ldots, \epsilon, |M, \ldots, M, |M, \ldots, M)$$
$$q_{2n+2m+3} = (0, 0, |0, 0, |1, \ldots, 1, |0, \ldots, 0, |0, \ldots, 0);$$
$$p_{2n+2m+3} = (\epsilon, \epsilon, |n - 1, n - 1, |1, \ldots, 1, |\epsilon, \ldots, \epsilon, |\epsilon, \ldots, \epsilon).$$

We have now described the dataset S. Before embarking further on the proof, let us describe the main idea. Consider the n goods, $5, 6, 7, \ldots, n + 4$ in observation $2n + 2m + 3$. Each of these goods corresponds to a variable in the instance of MNAE-3SAT. We will argue that each of these n goods is allocated for a large part (i.e. $\geq \frac{3}{4}$) to some member $\ell \in \{1, 2\}$. This is akin to setting the corresponding variable to TRUE (if the good goes for the larger part to member 1), or to FALSE (if the good goes for the larger part to member 2). Of course it remains to show that this is a satisfying truth assignment.

According to Definition 1, recall that we say that for member $\ell \in \{1, 2\}$ q_a^ℓ is directly revealed preferred to q_b^ℓ, when we have $p_a q_a^\ell \geq p_a q_b^\ell$ with $a, b \in S$.

Claim 1. *If $p_a q_a \geq p_a q_b$ for some $a, b \in S$, $a \neq b$, then there exists an $\ell \in \{1, 2\}$ for which q_a^ℓ is directly revealed preferred to q_b^ℓ.*

Proof. Consider any split of q_a into q_a^1, q_a^2, and q_b into q_b^1, q_b^2, i.e., let $q_a^1 + q_a^2 = q_a$ and $q_b^1 + q_b^2 = q_b$. Since $p_a q_a \geq p_a q_b$, it follows that

[2] In some datasets we have some information on how consumption is shared. This is called assignable information or exclusive goods in the literature. Such information can easily be integrated in our analysis and would not change our results.

$p_a(q_a^1+q_a^2) \geq p_a(q_b^1+q_b^2)$. Hence, either $p_a q_a^1 \geq p_a q_b^1$ or $p_a q_a^2 \geq p_a q_b^2$ (or both). □

Notice that, apart from bundle $q_{2n+2m+3}$, all other bundles are unit vectors. We will use $q_{i,j}$ ($p_{i,j}$) to denote the quantity (price) of good j in observation i, $i = 1, \ldots, 2n+2m+3$, $j = 1, \ldots, 3n+4$. We now exhibit a trick that we will use throughout the proof. Consider a hypothetical dataset, containing the observations a and b as follows:

$$q_a = (1, 0), \qquad p_a = (1, \epsilon)$$
$$q_b = (0, 1), \qquad p_b = (\epsilon, 1).$$

We say that the split of a bundle is *extreme* if each unit good of that bundle goes to one (of the two) members with fraction at least $1 - \epsilon$.

Claim 2. *In any feasible solution to private* 2-WARP *of some dataset containing observations a and b, the split of bundles q_a and q_b is extreme.*

Proof. Clearly, we have both $p_a q_a > p_a q_b$ and $p_b q_b > p_b q_a$. So using Claim 1, it follows that for one member ℓ we have that q_a^ℓ is directly revealed preferred to q_b^ℓ and simultaneously for one member ℓ' $q_b^{\ell'}$ is directly revealed preferred to $q_a^{\ell'}$. Thus in any feasible solution ℓ and ℓ' must be different (otherwise private 2-WARP is violated). Let us assume, without loss of generality, that for member 1 we have that q_a^1 is directly revealed preferred to q_b^1 and q_b^1 is not directly revealed preferred to q_a^1. Let α be the fraction of bundle a allocated to member 1, and β the fraction of bundle b allocated to this member. We then have that

$$p_b q_b^1 < p_b q_a^1 \Rightarrow \beta < \epsilon\alpha.$$

Since $\alpha \leq 1$, we conclude $\beta < \epsilon = \frac{1}{4n}$. Likewise, since for member 2 we have that q_a^2 is not directly revealed preferred to q_b^2, we find:

$$p_a q_a^2 < p_a q_b^2 \Rightarrow 1 - \alpha < \epsilon(1-\beta) \Rightarrow \alpha > 1 - \epsilon = \frac{4n-1}{4n}.$$

Claim 2 follows. □

Clearly, Claim 2 is applicable to any pair of observations involving bundles that are unit vectors, and price vectors that feature price ϵ and price 1. When applying this claim further on in our proof, some price vectors also include price M. However, whenever this is the case, the quantity of the goods will be 0 in both bundles, so the M prices can be ignored in these situations.

We now proceed to show that when the constructed instance of private 2-WARP is a YES-instances, a satisfying truth assignment exists.

Claim 3. *In any feasible solution to this instance of private* 2-WARP*, we have for $\ell = 1, 2$: $p_{2n+2m+3} q_{2n+2m+3}^\ell > 1$.*

Proof. Observe that Claim 2 is applicable to observations $2n+2m+1$ and $2n+2m+2$. Thus the split of the bundles $q_{2n+2m+1}$ and $q_{2n+2m+2}$ is extreme. Let us assume, without loss of generality, that good 3 is allocated to member 1 with fraction at least $1-\epsilon$, while good 4 is allocated to member 2 with fraction $1-\epsilon$. Thus:

$$p_{2n+2m+1} q_{2n+2m+1}^1 \geq 1 - \epsilon = \frac{4n-1}{4n} \geq \frac{1}{4} = n\frac{1}{4n} = n\epsilon \geq p_{2n+2m+1} q_{2n+2m+3}. \quad (1)$$

It follows that for member 1, in any feasible solution, observation $q_{2n+2m+1}^1$ is revealed preferred over $q_{2n+2m+3}^1$. Then, in order to satisfy private 2-WARP, we must have:

$$p_{2n+2m+3} q_{2n+2m+3}^1 < p_{2n+2m+3} q_{2n+2m+1}^1 \leq p_{2n+2m+3} q_{2n+2m+1} \Rightarrow \quad (2)$$

$$\sum_{i=1}^{n} q_{2n+2m+3,4+i}^1 < n - 1. \quad (3)$$

Since, for $a \in S$, $q_a^2 = q_a - q_a^1$, we derive, using (3):

$$\sum_{i=1}^{n} q_{2n+2m+3,4+i}^2 = n - \sum_{i=1}^{n} q_{2n+2m+3,4+i}^1 > 1. \quad (4)$$

Finally, since $p_{2n+2m+3,i} = 1$ for $i = 5, 6, \ldots, n+4$, it follows that (4) can be written as:

$$p_{2n+2m+3} q_{2n+2m+3}^2 > 1.$$

A similar reasoning involving member 2 and observations $2n+2m+2$ and $2n+2m+3$ leads to:

$$p_{2n+2m+3} q_{2n+2m+3}^1 > 1. \quad □$$

To proceed, let us consider observation i, ($1 \leq i \leq n$), and observation $2n+2m+3$. Using Claim 3, we observe:

$$p_{2n+2m+3} q_{2n+2m+3}^\ell > 1 > p_{2n+2m+3} q_i \quad \text{for } \ell = 1, 2, 1 \leq i \leq n. \quad (5)$$

Thus, no matter the split of q_i into q_i^1 and q_i^2, for both member 1 and member 2 we have that $q_{2n+2m+3}^1$ (resp. $q_{2n+2m+3}^2$) is directly revealed preferred over observation q_i^1 (resp. q_i^2), with $i = 1, \ldots, n$. Since we have a YES-instance of 2-WARP, we know that then, for $\ell = 1, 2$:

$$p_i q_i^\ell < p_i q_{2n+2m+3}^\ell. \quad (6)$$

Observe that Claim 2 is applicable to observations i and $n+i$. Thus, the split of q_i and q_{n+i} is extreme. Hence, there is a member ℓ for which:

$$p_i q_i^\ell \geq 1 - \epsilon. \quad (7)$$

Inequalities (6) and (7) imply that the split of $q_{2n+2m+3}$ is such that:

$$p_i q_{2n+2m+3}^\ell > 1 - \epsilon. \quad (8)$$

Consider the vectors p_i and $q_{2n+2m+3}$, $1 \leq i \leq n$. It follows that:

$$p_i q_{2n+2m+3}^\ell = \epsilon \sum_{j=5, j\neq 4+i}^{n+4} q_{2n+2m+3,j}^\ell + q_{2n+2m+3,4+i}^\ell. \quad (9)$$

Also:

$$\sum_{j=5, j\neq 4+i}^{n+4} q_{2n+2m+3,j}^\ell \leq \sum_{j=5, j\neq 4+i}^{n+4} q_{2n+2m+3,j} = n - 1. \quad (10)$$

Rewriting (9), and using inequalities (8) and (10) gives for each $i = 1, \ldots, n$:

$$q_{2n+2m+3,4+i}^\ell = p_i q_{2n+2m+3}^\ell - \epsilon \sum_{j=5, j\neq 4+i}^{n+4} q_{2n+2m+3,j}^\ell > 1 - \epsilon - \epsilon(n-1) = 1 - \frac{n}{4n} = \frac{3}{4}. \quad (11)$$

Concluding, each good $i = 5, 6, \ldots, n+4$ in observation $2n+2m+3$ is allocated for over $\frac{3}{4}$ to some member $\ell \in \{1, 2\}$.

Finally, we look at the two observations corresponding to each clause $j = 1, \ldots, m$. It is clear that for each member ℓ we have that

$q^{\ell}_{2n+2m+3}$ is directly revealed preferred over both observations q^{ℓ}_{2n+j} and q^{ℓ}_{2n+m+j}. Observe also that Claim 2 is applicable to observations $2n + j$ and $2n + 2m + j$. Thus, in order not to have a violation of private 2-WARP, we should have for member ℓ q^{ℓ}_{2n+j} is not directly revealed preferred over $q^{\ell}_{2n+2m+3}$. Thus, for each $\ell = 1, 2$:

$$p_{2n+j}q^{\ell}_{2n+j} < p_{2n+j}q^{\ell}_{2n+2m+3}. \quad (12)$$

Since (without loss of generality), for member 1, we have $p_{2n+j}q^{1}_{2n+j} \geq 1 - \epsilon$, and thus we have using (12):

$$p_{2n+j}q^{1}_{2n+2m+3} > 1 - \epsilon. \quad (13)$$

This means that one of the three goods associated to clause j is allocated over $\frac{3}{4}$ to member 1. We argue by contradiction. Indeed, in case none of the three goods of clause j are allocated over $\frac{3}{4}$ to member 1, then they are allocated for at most $\frac{1}{4}$ to member 1. Then,

$$p_{2n+j}q^{1}_{2n+2m+3} \leq 3\frac{1}{4} + (n-3)\epsilon = \frac{3}{4} + \frac{n-3}{4n} = \frac{4n-3}{4n} < \frac{4n-1}{4n}. \quad (14)$$

Thus we would have $p_{2n+j}q^{1}_{2n+2m+3} < 1 - \epsilon$, contradicting (13). Therefore, at least one of the goods associated with j is allocated over $\frac{3}{4}$ to member 1. Clearly, a similar reasoning involving $2n + 2m + j$ and member 2 implies that one of these three goods must be allocated over $\frac{3}{4}$ to member 2.

In conclusion, we now know the following about any valid allocation of observation $2n + 2m + 3$ which satisfies private 2-WARP. Firstly, that each good is split up in a large and a small allocation for the different members. Secondly, that for each clause and each member, there is at least one of the goods associated with the variables that has a large allocation. A valid truth assignment for MNAE 3-SAT can now be found as follows. If, in observations $2n+2m+3$ a good is largely allocated to member 1, the variable is set to TRUE, if a good is largely allocated to member 2, the variable is FALSE.

If we have a Yes-instance of MNAE 3-SAT, an allocation of goods which satisfies private 2-WARP exists. For observation $2n+2m+3$, fully assign each good associated with a TRUE variable to member 1, and each good associated with a FALSE variable to member 2. Likewise, fully assign the bundle i to member 1 if x_i is TRUE and to member 2 if it is FALSE. Furthermore, for all $j = 1, \ldots, m$, fully assign bundles $2n+j$ to member 1 and all $2n+m+j$ to member 2. Finally, fully assign $2n + 2m + 1$ to member 1 and $2n + 2m + 2$ to member 2. It can be easily checked that such an allocation satisfies private 2-WARP. □

The following argument generalizes our NP-completeness results for private 2-WARP towards private k-WARP for any fixed $k \geq 2$.[3]

Theorem 2. *Testing private k-WARP is NP-complete for any fixed $k \geq 2$.*

Proof. NP-Completeness for private k-WARP, $k > 2$ can also be proven through a reduction from MONOTONE NOT-ALL-EQUAL 3-SAT. We briefly sketch this reduction. The dataset constructed is the same as for $k = 2$, except that for any additional member beyond the second, one extra observation and one extra good is added. There are now $2n+2m+3+(k-2)$ observations and $3n+4+(k-2)$ goods. Observation $2n + 2m + 3 + i$ consists of only one unit of good $3n + 4 + i$, which has price 1 and all other prices equal to ϵ. In all other observations, good $3n + 4 + i$ has price ϵ. Using an argument similar to the proof of Claim 2, it is clear that, in any feasible solution, allocations are extreme, i.e., any member ℓ who is allocated more than a small fraction of good $3n+4+i$ will prefer $q^{\ell}_{2n+2m+3+i}$ over all other bundles and any member who is allocated more than a fraction of a bundle in other observations prefers that bundle over $q_{2n+2m+3+i}$. Any feasible split will thus have the goods $3n + 4 + i$ in the extra observations almost completely allocated to the members $3, 4, \ldots,$ while all of the bundles present in the proof for $k = 2$ must still be split over two members. □

4. The collective consumption model with only public goods

We next turn to the collective model with all goods publicly consumed in the household. In this case all member-specific utility functions are all defined for the same bundle of public goods. For ease of exposition, we again mainly focus on households consisting of two members. But, like before, our findings for this case are easily extended to k-member households (with $k \geq 2$).

For this public setting, we formalize the idea that household members have individual-specific (unobserved) willingness-to-pay for the public goods (bought at prices p_t). To do so, for each observation $t \in N$, we define a *feasible personalized price vector* (p^1_t, p^2_t). Intuitively, these feasible personalized prices capture the fractions of the household prices for the public goods that are borne by the individual members ℓ. Given the Pareto efficiency assumption that underlies the collective consumption model, these prices can also be interpreted as Lindahl prices. We refer to Cherchye et al. (2011) for a detailed discussion.

Similar to before, for each member ℓ ($\ell = 1, 2$) we consider *personalized consumption datasets* by $S_\ell = \{(p^{\ell}_t, q_t) : t \in N\}$. The extension of WARP to this collective consumption model is then as follows.

Definition 3 (*Public 2*-WARP)**.** Let $S = \{(p_t, q_t) : t \in N\}$ be a dataset of a two-member household. We say that S is consistent with *public* 2-WARP if and only if:

(i) For each $t \in N$ there exist $p^1_t, p^2_t \in \mathbb{R}^m_+$ such that $p_t = p^1_t + p^2_t$, and
(ii) For each member $\ell \in \{1, 2\}$, the set $S_\ell = \{(p^{\ell}_t, q_t) : t \in N\}$ satisfies WARP.

This problem can be rephrased as the following decision problem:

Problem: public 2-WARP

Instance: A dataset $S = \{(p_t, q_t) : t \in N\}$.

Question: Do there exist $p^1_t, p^2_t \in \mathbb{R}^m_+$ satisfying $p_t = p^1_t + p^2_t$ for each $t \in N$ such that for $\ell = 1, 2$, the set $S_\ell = \{(p^{\ell}_t, q_t) : t \in N\}$ satisfies WARP?

It turns out that answering this question also implies solving an NP-complete problem, even for two members in the household.

Theorem 3. *Testing public* 2-WARP *is NP-complete.*

Proof. We again use a reduction from MONOTONE NOT-ALL-EQUAL 3-SAT.

Instance: A set of variables $X = \{x_1, x_2, \ldots, x_n\}$ and a set of clauses $C = \{c_1, c_2, \ldots, c_m\}$ with each clause consisting of 3 non-negated literals.

Question: Does there exist a truth-assignment so that for each clause, either one or two of the literals are TRUE?

It is not difficult to see that public 2-WARP belongs to the class NP. The rest of this proof is structured as follows: given an arbitrarily instance of MNAE 3-SAT, we first build an instance of

[3] The definition of private k-WARP is trivially analogous to the one of private 2-WARP. For compactness, we do not include it here.

public 2-WARP and next, we prove that we have a YES instance of MNAE 3-SAT if and only if the constructed instance of public 2-WARP is a YES instance.

Consider an arbitrary instance $X = \{x_1, x_2, \ldots, x_n\}$ and $C = \{c_1, c_2, \ldots, c_m\}$ of MNAE 3-SAT. We build an instance of public 2-WARP using $3n + 2$ goods and $2n + m + 2$ observations. We next describe the quantity and the price of the goods for each observation. We use $\epsilon = \frac{1}{4n}$ and $M = n + 1$. The first block of $2n$ observations corresponds to the variables and is given by:

$$q_1 = (2, 0, \ldots, 0, |0, \ldots, 0, |1, 0, \ldots, 0, |0, 0);$$
$$p_1 = (1, \epsilon, \ldots, \epsilon, |\epsilon, \ldots, \epsilon, |\epsilon, M, \ldots, M, |M, M)$$
$$q_2 = (0, 2, \ldots, 0, |0, \ldots, 0, |0, 1, \ldots, 0, |0, 0);$$
$$\vdots$$
$$q_n = (0, 0, \ldots, 2, |0, \ldots, 0, |0, 0, \ldots, 1, |0, 0);$$
$$p_2 = (\epsilon, 1, \ldots, \epsilon, |\epsilon, \ldots, \epsilon, |M, \epsilon, \ldots, M, |M, M)$$
$$\vdots$$
$$p_n = (\epsilon, \epsilon, \ldots, 1, |\epsilon, \ldots, \epsilon, |M, M, \ldots, \epsilon, |M, M)$$
$$q_{n+1} = (0, \ldots, 0, |1, 0, \ldots, 0, |1, 0, \ldots, 0, |0, 0);$$
$$\vdots$$
$$q_{2n} = (0, \ldots, 0, |0, 0, \ldots, 1, |0, 0, \ldots, 1, |0, 0);$$
$$p_{n+1} = (\epsilon, \ldots, \epsilon, |1, \epsilon, \ldots, \epsilon, |\epsilon, M, \ldots, M, |M, M)$$
$$\vdots$$
$$p_{2n} = (\epsilon, \ldots, \epsilon, |\epsilon, \epsilon, \ldots, 1, |M, M, \ldots, \epsilon, |M, M)$$

Notice that each entry i in both a price-vector and a quantity-vector correspond to good i, $i = 1, \ldots, 3n + 2$. The second block of m observations corresponds to the clauses. For each clause $c_a = \{x_i, x_j, x_k\}$, we have the observation $2n + a$ $(a = 1, \ldots, m)$.

$$q_{2n+1} = (\{0, 1\}|0, \ldots, 0|0, \ldots, 0|0, 0);$$
$$p_{2n+1} = (\{M, 1\}|M, \ldots, M|\epsilon, \ldots, \epsilon|2, 2)$$
$$q_{2n+2} = (\{0, 1\}|0, \ldots, 0|0, \ldots, 0|0, 0);$$
$$\vdots$$
$$q_{2n+m} = (\{0, 1\}|0, \ldots, 0|0, \ldots, 0|0, 0);$$
$$p_{2n+2} = (\{M, 1\}|M, \ldots, M|\epsilon, \ldots, \epsilon|2, 2)$$
$$\vdots$$
$$p_{2n+m} = (\{M, 1\}|M, \ldots, M|\epsilon, \ldots, \epsilon|2, 2)$$

In the observations corresponding to a given clause c_a, $a = 1, \ldots, m$, the quantity of good i is 1 if the variable x_i is part of the clause c_a, and 0 otherwise. As for the prices, the price of good i is 1 if variable x_i is in clause c_a, M otherwise. Finally, we have observations $2n + m + 1$ and $2n + m + 2$:

$$q_{2n+m+1} = (0, \ldots, 0, |0, \ldots, 0, |0, \ldots, 0|1, 0);$$
$$p_{2n+m+1} = (\epsilon, \ldots, \epsilon, |\epsilon, \ldots, \epsilon, |M, \ldots, M|1, \epsilon)$$
$$q_{2n+m+2} = (0, \ldots, 0, |0, \ldots, 0, |0, \ldots, 0|0, 1);$$
$$p_{2n+m+2} = (\epsilon, \ldots, \epsilon, |\epsilon, \ldots, \epsilon, |M, \ldots, M|\epsilon, 1).$$

This concludes the description of the instance of public 2-WARP. The main idea used to argue the equivalence between MONOTONE NOT-ALL-EQUAL 3-SAT and public 2-WARP is as follows. The first n goods, $1, \ldots, n$ represent the variables considered. We will argue that the (unit) prices of each of these n goods in these first n observations are allocated for a large part (i.e. $\geq 1 - \epsilon$) to some member $\ell \in \{1, 2\}$. This corresponds to setting the variable to TRUE (if the price goes for the larger part to member 1), or to FALSE (if the price goes for the larger part to member 2). The proof will show that this is a satisfying truth assignment.

We first show that the trick used in the proof for private goods can be used similarly in the context of public goods. Consider a hypothetical dataset, containing the observations a and b as follows:

$$q_a = (1, 0), \qquad p_a = (1, \epsilon)$$
$$q_b = (0, 1), \qquad p_b = (\epsilon, 1).$$

We say that the split of the prices of a bundle is *extreme* if the unit prices of goods present in the bundle are allocated to one of the two members with amount at least $1 - \epsilon$.

Claim 4. *In any feasible solution to public 2-WARP of some dataset containing observations a and b as above, the split of the price vector is extreme.*

Proof. Clearly, we have both $p_a q_a > p_a q_b$ and $p_b q_b > p_b q_a$. In this case, we have for one member ℓ q_a is directly revealed preferred over q_b, and for one member ℓ' q_b is directly revealed preferred over q_a. Thus in any feasible solution ℓ and ℓ' must be different (otherwise public 2-WARP is violated). Let us assume, without loss of generality, that for member 1 q_b is directly revealed preferred over q_a, while q_a is not directly revealed preferred over q_b. Let α be the part of the price of product 1 in bundle a allocated to member 1, and β the part of the price of product 2 allocated to this member. We then have

$$p_a^1 q_a < p_a^1 q_b \Rightarrow \alpha < \beta.$$

Since $\beta \leq \epsilon$, we conclude $\alpha < \epsilon = \frac{1}{4n}$. This implies the price of the first good of a is allocated for more than $1 - \epsilon$ to member 2. By the same argument, the price of the second good in observation b is allocated for more than $1 - \epsilon$ to member 1. Claim 4 follows. □

Claim 5. *In any feasible solution, the decomposition of p_{2n+m+1} is such that for some member ℓ we have that q_{2n+m+1} is directly revealed preferred over all bundles q_{2n+i}, with $i = 1, \ldots, m$. While for the other member $\ell' \neq \ell$ q_{2n+m+2} is directly revealed preferred over all q_{2n+i}.*

Proof. It is clear that Claim 4 may be directly applied to observations $2n + m + 1$ and $2n + m + 2$. Without loss of generality, we assume that in any feasible solution, the price of good $3n+1$ in observation $2n + m + 1$ is allocated almost completely to member 1, while the price of good $3n+2$ in observations $2n+m+2$ is allocated to member 2. It can easily be checked that $p^1_{2n+m+1} q_{2n+m+1} > 1-\epsilon$, while $p^1_{2n+m+1} q_{2n+i} \leq 3\epsilon$ for all $i = 1, \ldots, m$. A similar analysis for member 2 proves the claim. □

This claim allows us to find a condition on the split of the prices in the observations associated with the clauses.

Claim 6. *In any feasible solution, for any observation $2n + i$ and any member ℓ, the split of the corresponding prices must be so that $1 < p^\ell_{2n+i} q_{2n+i} < 2$.*

Proof. By Claim 5, for any member ℓ, we have that either q_{2n+m+1} is directly revealed preferred to q_{2n+i} or q_{2n+m+2} is directly revealed preferred to q_{2n+i}. Without loss of generality, we assume q_{2n+m+1} is directly revealed preferred to q_{2n+i}. In any feasible solution, it is then the case that $p^\ell_{2n+i} q_{2n+i} < p^\ell_{2n+i} q_{2n+m+1}$. It can be easily checked that $p^\ell_{2n+i} q_{2n+m+1}$ is at most 2. As $p_{2n+i} q_{2n+i} = 3$ and $p^1_{2n+i} + p^2_{2n+i} = p_{2n+i}$, $p^\ell_{2n+i} q_{2n+i} > 1$ follows immediately. □

Consider now a pair of observations i and $n + i$, $i = 1, \ldots, n$. While these prices and quantities do not coincide with those in Claim 4 exactly, it can be easily seen that the split of price i in observation i is extreme. Next, notice that the member ℓ to whom more than $1 - \epsilon$ of the price of i is allocated, will have $p^\ell_i q_i > p^\ell_i q_{2n+a}$, with $a = 1, \ldots, m$. This brings us to the following claim.

Claim 7. *In any feasible solution to public 2-WARP, if there exists some clause c_a, $a = 1, \ldots, m$ with variables x_i, x_j, x_k, it cannot be the case that for some member ℓ, the prices of i, j and k in respectively observations i, j and k are allocated for more than $1 - \epsilon$ to ℓ.*

Proof. We argue by contradiction. Suppose that in a feasible solution to public 2-WARP some member ℓ is allocated almost completely the prices of goods i, j, k occurring in some clause c_a in respective observations i, j, k. The following inequalities follow: $p_i^\ell q_i > p_i^\ell q_{2n+a}, p_i^\ell q_i > p_i^\ell q_{2n+a}, p_i^\ell q_i > p_i^\ell q_{2n+a}$. Hence, in order not to violate WARP for member ℓ, we must have:

$$p_{2n+a}^\ell q_{2n+a} < p_{2n+a}^\ell q_i \tag{15}$$

$$p_{2n+a}^\ell q_{2n+a} < p_{2n+a}^\ell q_j \tag{16}$$

$$p_{2n+a}^\ell q_{2n+a} < p_{2n+a}^\ell q_k. \tag{17}$$

As Claim 6 shows that $p_{2n+c}^\ell q_{2n+c} > 1$, we must have $p_{2n+c}^\ell q_i > 1$. This can be rewritten as $2 \times p_{2n+c,i}^\ell + p_{2n+c,2n+i}^\ell > 1$, for convenience, we will ignore $p_{2n+c,2n+i}^\ell$ as it is negligible in the following analysis. We now have $p_{2n+c,i}^\ell > \frac{1}{2}$. However, this must hold for all three prices associated with i, j, k, which gives the following $p_{2n+c}^\ell q_{2n+c} = p_{2n+c,i}^\ell + p_{2n+c,j}^\ell + p_{2n+c,k}^\ell > \frac{3}{2}$. In this case, $p_{2n+c,i}^\ell > \frac{1}{2}$ is no longer sufficient, as $2 \times p_{2n+c,i}^\ell + p_{2n+c,2n+i}^\ell > \frac{3}{2}$ is required. By the same argument, we obtain that $p_{2n+c,i}^\ell > \frac{3}{4}$. However, in this case $p_{2n+c,i}^\ell + p_{2n+c,j}^\ell + p_{2n+c,k}^\ell > \frac{9}{4}$. However, by Claim 6 this cannot be the case for a feasible solution. By contradiction, Claim 7 is thus proven. □

Now, it has become easy to show that a YES-instance of public 2-WARP problem corresponds to a satisfying truth assignment in MNAE- 3-SAT, and vice versa. It is clear that – if a feasible solution to public 2-WARP exists – the prices of good i in observation i, $i = 1, \ldots, n$ are always allocated with an extreme split. If the price of good i in observation i is almost completely allocated to member 1, we set the corresponding variable x_i to TRUE, otherwise we set it to FALSE. By Claim 7, we know that if a feasible solution to public WARP exists, and goods i, j, k are in a clause, no member will have the prices of all 3 goods allocated to him/her. Thus, a solution to public 2-WARP corresponds to a satisfying truth assignment in MNAE- 3-SAT. The other direction, i.e., finding a solution to public 2-WARP when a satisfying truth assignment in MNAE- 3-SAT is given is easy: simply allocate almost completely good i in observation i to member 1 if x_i is TRUE, else allocate good i almost completely to member 2, $i = 1, \ldots, n$. For observations $n + i$ and goods $n + 1$, the reverse is done. $2n + m + 1$ and $2n + m + 2$ are respectively allocated to members 1 and 2. All other prices may be split evenly between the members. This will satisfy public 2-WARP. □

Similar to the private setting, we can extend our NP-completeness results for public 2-WARP to public k-WARP for any fixed $k \geq 2$.[4]

Theorem 4. *Testing public k-WARP is NP-complete for any fixed $k \geq 2$.*

Proof. The proof for this Theorem is analogous to the proof of Theorem 2. □

[4] Again, the definition of public k-WARP is directly analogous to the one of public 2-WARP and, therefore, we do not include it here.

5. The general collective consumption model

The final collective consumption model that we consider is the most general one. It does not make any assumption regarding the nature of the consumed goods. That is, every good can be privately or publicly consumed, and the private goods may generate externalities. If a dataset S is a YES-instance to either public or private 2-WARP, it is thus also a YES-instance to general 2-WARP. Clearly, the converse is not necessarily true. As before, we only observe data at the aggregate household level.

Different from before, we no longer use the notions of feasible personalized prices and quantities to characterize this general collective model. Instead, we follow the approach developed in Cherchye et al. (2007, 2012), which defines a revealed preference characterization in terms of *hypothetical preference relations*. More precisely, for some member ℓ, we denote by H_0^ℓ the hypothetical preference of that member. The expression "$q_s H_0^\ell q_t$" means that we hypothesize that member ℓ *directly prefers* the bundle q_s over the bundle q_t (for $s, t \in N$). In Cherchye et al. (2007) these hypothetical relations are then used to derive necessary conditions that the data need to satisfy in order to be compatible with the general collective consumption model. We refer to Cherchye et al. (2007, 2012) for a detailed discussion.

Given our specific objective, we consider an extension of WARP to 2-member households that makes use of this notion of hypothetical preferences. This extension is derived from the revealed preference characterization in Proposition 2 of Cherchye et al. (2007), by essentially dropping the transitivity requirement.

Definition 4 (*General 2*-WARP). Let $S = \{(p_t, q_t) : t \in N\}$ be a dataset of a two-member household. We say that S is consistent with *general* 2-WARP if and only if there exist hypothetical preferences H_0^1, H_0^2 that satisfy:

(a) For each pair of distinct observations $s, t \in N$: if $p_s q_s \geq p_s q_t$, then $q_s H_0^1 q_t$ or $q_s H_0^2 q_t$;
(b) For each pair of distinct observations $s, t \in N$: if $p_s q_s \geq p_s q_t$, $q_t H_0^\ell q_s$ and $q_t \neq q_s$, then $q_s H_0^r q_t$, with $\ell, r \in \{1, 2\}$ and $\ell \neq r$;
(c) For each three distinct observations $s, t, u \in N$: if $p_s q_s \geq p_s (q_t + q_u)$ and $q_t H_0^\ell q_s$, then $q_s H_0^r q_u$ with $\ell, r \in \{1, 2\}$ and $\ell \neq r$;
(d) For each pair of distinct observations $s, t \in N$: if $q_s H_0^1 q_t$, $q_s H_0^2 q_t$ and $q_t \neq q_s$, then $p_t q_t < p_t q_s$;
(e) For each three distinct observations $s, t, u \in N$: if $q_s H_0^1 q_t$ and $q_u H_0^2 q_t$, then $p_t q_t < p_t (q_s + q_u)$.

In what follows, we will show that it is possible to check this general 2-WARP condition efficiently (i.e. in polynomial time), which contrasts with our results for the private and public settings in the previous sections. Importantly, we will only show this complexity result for the two-member case. Different from before, the result is not straightforwardly generalized towards the general case with k household members ($k \geq 2$). We leave the study of this k-member case for future research.[5]

As a first step towards formulating the decision problem corresponding to our definition of general 2-WARP, we define a simplification of the above definition that is easier to use. Specifically, we replace condition (b) with a closely similar, but somewhat more stringent condition, and we drop conditions (d) and (e).

[5] In this respect, we note that existing applications of the collective model usually consider households with only two decision makers (e.g. husband and wife, with expenses on children treated as public consumption). Therefore, we may safely argue that the two-member case is the most relevant one from a practical perspective.

Definition 5 (*General 2-WARP*). Let $S = \{(p_t, q_t) : t \in N\}$ be a dataset of a two-member household. We say that S is consistent with *general* 2-WARP if and only if there exist hypothetical preferences H_0^1, H_0^2 that satisfy:

(i) For each pair of distinct observations $s, t \in N$: if $p_s q_s \geq p_s q_t$, then $q_s H_0^1 q_t$ or $q_s H_0^2 q_t$;
(ii) For each pair of distinct observations $s, t \in N$: if $p_s q_s \geq p_s q_t$, $q_t H_0^\ell q_s$ and $q_t \neq q_s$, then $\neg(q_s H_0^\ell q_t)$, with $\ell \in \{1, 2\}$;
(iii) For each three distinct observations $s, t, u \in N$: if $p_s q_s \geq p_s(q_t + q_u)$ and $q_t H_0^\ell q_s$, then $q_s H_0^r q_u$ with $\ell, r \in \{1, 2\}$ and $\ell \neq r$;

Claim 8. *Definitions 4 and 5 are equivalent.*

Proof. ($\Rightarrow$) Assume that there exist hypothetical relations for which conditions (a)–(e) in Definition 4 are satisfied. It is clear that conditions (i) and (iii) are then also satisfied, as these are identical to (a) and (c). Now suppose the hypothetical relations include a violation of (ii), i.e., there exist distinct observations $s, t \in N$, for which $p_s q_s \geq p_s q_t$, $q_t H_0^1 q_s$, $q_s H_0^1 q_t$ and $q_t \neq q_s$. We then have to consider two scenarios: either $p_t q_t \geq p_t q_s$ or $p_t q_t < p_t q_s$. If $p_t q_t < p_t q_s$, then there is no need to specify $q_t H_0^1 q_s$ and thus condition (ii) is by construction satisfied. In the alternative scenario, $p_t q_t \geq p_t q_s$, then $q_s H_0^1 q_t$ implies that $q_t H_0^2 q_s$ (since condition (b) is satisfied). But this entails a violation of condition (d), since $p_s q_s \geq p_s q_t$, $q_t H_0^1 q_s$ and $q_t H_0^2 q_s$. This gives us the desired contradiction.

($\Leftarrow$) Next assume there exist hypothetical relations for which Conditions (i)–(iii) in Definition 5 are satisfied. Again, conditions (a) and (c) are identical to (i) and (iii) and are satisfied. Next, if (b) is violated, there exist $s, t \in N$ such that $p_s q_s \geq p_s q_t$, $q_t H_0^1 q_s$, $q_t \neq q_s$, and $\neg(q_s H_0^2 q_t)$. Since condition (i) requires either $q_s H_0^1 q_t$ or $q_s H_0^2 q_t$, it must be the case that $q_s H_0^1 q_t$, which violates (ii), thus there is violation of (b) if (i)–(iii) are satisfied. Now suppose (d) is violated. Then $q_s H_0^1 q_t$, $q_s H_0^2 q_t$ and $p_t q_t \geq p_t q_s$. By rule (i), it must then be the case that $q_t H_0^1 q_q$ or $q_t H_0^2 q_q$, either of which again violates (ii). Finally, a violation of (e) implies $q_s H_0^1 q_t$, $q_u H_0^2 q_t$ and $p_t q_t \geq p_t(q_s + q_u)$. To satisfy condition (iii), if $p_t q_t \geq p_t(q_s + q_u)$ and $q_s H_0^1 q_t$, it must be the case that $q_t H_0^2 q_u$, since we also have $q_u H_0^2 q_t$, either (ii) or (iii) must be violated if (e) is violated. □

The problem of testing whether a collective rationalization of S exists is then formulated as the following *decision problem*:

Problem 2-WARP.

Instance: A dataset $S := \{(p_t, q_t) : t \in N\}$.

Question: Do there exist hypothetical preferences H_0^1, H_0^2, such that conditions (i)–(iii) in Definition 5 hold?

Before studying this decision problem more in detail, we want to make the following remarks. If the dataset S contains only three observations, let us say s, t, and u, then the answer to the decision problem is No if and only if the following three inequalities hold: $p_s q_s \geq p_s(q_t + q_u)$, $p_t q_t \geq p_t(q_s + q_u)$, and $p_u q_u \geq p_u(q_s + q_t)$. For datasets containing more than three observations, however, the presence of these three inequalities is not necessary to have a No answer. Indeed, the reader can check that the following inequalities involving four observations, let us say s, t, u, and v, also lead to a No answer to 2-WARP: $p_s q_s \geq p_s q_t$, $p_t q_t \geq p_t(q_s + q_u)$, $p_t q_t \geq p_t(q_s + q_v)$, $p_u q_u \geq p_u(q_t + q_v)$, and $p_v q_v \geq p_v(q_t + q_u)$. Furthermore, we mention that if there is no inequality of the form $p_s q_s \geq p_s(q_t + q_u)$ for all triples s, t, and u in N then we have a Yes instance of 2-WARP.

5.1. A graph interpretation of 2-WARP

We translate conditions (i) to (iii) into a directed graph setting (see Talla Nobibon et al., 2011 for a related construction). We build a directed graph $G = (V, A)$ from the dataset $S := \{(p_t, q_t) : t \in N\}$ as follows. A pair of distinct observations (s, t) with $s, t \in N$ represents a vertex in V if and only if both $p_s q_s \geq p_s q_t$ and $p_t q_t \geq p_t q_s$. Notice that V contains $O(n^2)$ vertices and if the vertex (s, t) exists then the vertex (t, s) also exists. The set of arcs A is defined in two steps as follows:

1: First, there is an arc from a vertex (s, t) to a vertex (u, v) whenever $t = u$.
2: Second, for any three distinct observations $s, t, u \in N$ satisfying $p_s q_s \geq p_s(q_t + q_u)$, $p_t q_t \geq p_t q_s$, $p_u q_u \geq p_u q_s$, we have an arc from (s, u) to (t, s), and from (s, t) to (u, s).

Notice that Step 1 ensures that there is an arc from node (s, t) to node (t, s) and vice versa. This graph construction differs from the one used when checking whether a dataset of a unitary household satisfies WARP: in that case, a directed graph is built where a vertex corresponds with an observation and there is an arc from s to t if and only if $p_s q_s \geq p_s q_t$. That approach is not considered because it is not quite clear how to deal with inequalities of the form $p_s q_s \geq p_s(q_t + q_u)$.

Given the directed graph $G = (V, A)$ built above, we define the 2-*undirected graph* $G_2 = (V, E)$ associated with G as the undirected graph obtained from G by transforming any pair of arcs forming a cycle of length 2 into a single edge (undirected arc); more precisely, $\{v_1, v_2\} \in E$ if and only if $v_1 v_2 \in A$ and $v_2 v_1 \in A$.

As an illustration of the graph construction, consider a dataset with three observations satisfying: $p_1 q_1 \geq p_1(q_2 + q_3)$, $p_2 q_2 \geq p_2(q_1 + q_3)$, $p_3 q_3 \geq p_3(q_1 + q_2)$, $p_2 q_2 \geq p_2 q_1$, and $p_3 q_3 \geq p_3 q_1$. This implies the existence of the vertices depicted in Fig. 1(a). The arcs stemming from Step 1 appear in Fig. 1(b), and the final graph is depicted in Fig. 1(c), where the dashed arcs are derived from Step 2. Finally, the 2-undirected graph G_2 associated with G is depicted in Fig. 1(d). We have the following result.

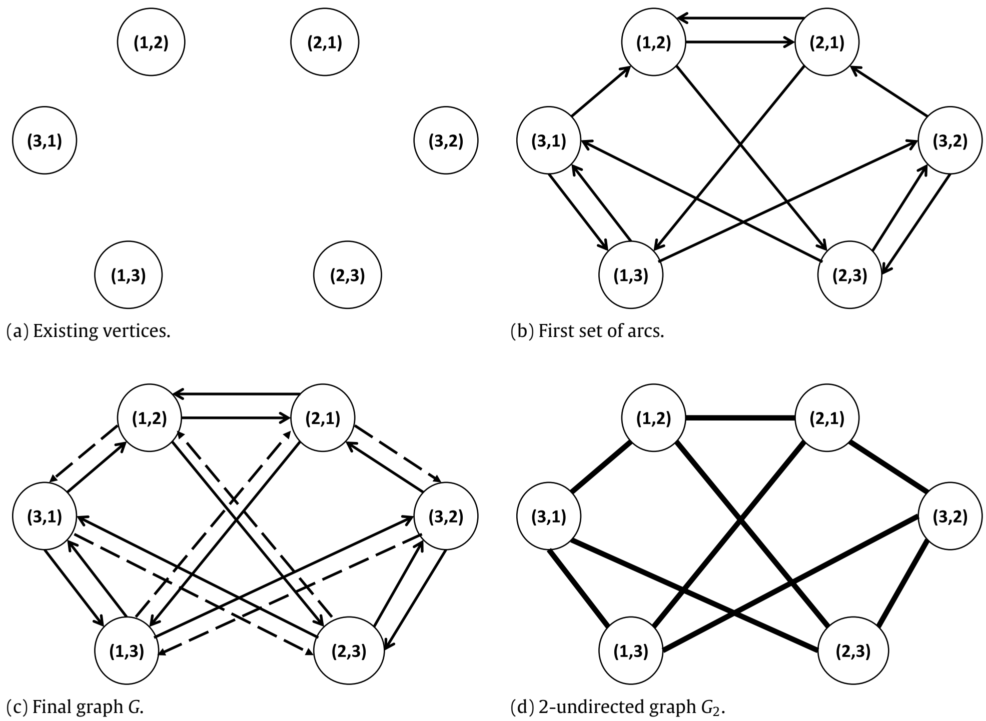
Theorem 5. *S is a* Y *es instance of* 2-WARP *if and only if the* 2-undirected *graph G_2 associated with G is bipartite.*

Proof. ($\Leftarrow$) Suppose that G_2 is bipartite. Thus, the set of vertices V can be partitioned into two subsets V_1 and V_2 such that each subset induces an independent set. In other words, $V = V_1 \cup V_2$, $V_1 \cap V_2 = \emptyset$, and there is no edge between two vertices of V_1 and no edge between two vertices of V_2. We build the hypothetical preferences H_0^1 and H_0^2 as follows: for every vertex $(s, t) \in V_1$ (respectively $(s, t) \in V_2$) we have $q_s H_0^1 q_t$ (respectively $q_s H_0^2 q_t$). Furthermore, for two distinct observations s and t such that $p_s q_s \geq p_s q_t$ and $(s, t) \notin V$, we set $q_s H_0^1 q_t$ and $q_s H_0^2 q_t$. This completes the definition of H_0^1 and H_0^2. Notice that there is no distinct pair of observations s, t for which we set $q_s H_0^\ell q_t$ and $q_t H_0^\ell q_s$ for some $\ell \in \{1, 2\}$. We now argue that H_0^1 and H_0^2 satisfy conditions (i) to (iii).

Condition (i): Let $s, t \in N$ be two distinct observations such that $p_s q_s \geq p_s q_t$. On the one hand, if $(s, t) \notin V$ then, by construction, $q_s H_0^1 q_t$ and $q_s H_0^2 q_t$. On the other hand, if $(s, t) \in V = V_1 \cup V_2$ then $(s, t) \in V_1$ or $(s, t) \in V_2$, and hence $q_s H_0^1 q_t$ or $q_s H_0^2 q_t$. Thus condition (i) is satisfied.

Condition (ii): As described above, there is no distinct pair of observations $s, t \in N$ for which we set $q_s H_0^\ell q_t$ and $q_t H_0^\ell q_s$ for some $\ell \in \{1, 2\}$. Thus condition (ii) is satisfied.

Condition (iii): Let $s, t, u \in N$ be three distinct observations such that $p_s q_s \geq p_s(q_t + q_u)$ and $q_t H_0^1 q_s$. There are two cases: (1) if $p_u q_u < p_u q_s$ then $(s, u) \notin V$ and, since $p_s q_s \geq p_s q_u$, we have by



(a) Existing vertices. (b) First set of arcs.

(c) Final graph G. (d) 2-undirected graph G_2.

Fig. 1. Illustration of the construction of G and the associated 2-undirected graph G_2.

construction of H_0^ℓ, $q_s\,H_0^1\,q_u$ and $q_s\,H_0^2\,q_u$, and we are done; (2) if $p_uq_u \geq p_uq_s$ then $(s,u) \in V$. Let us now argue by contradiction that $(t,s) \in V$. Indeed, if $(t,s) \notin V$, then $(s,t) \notin V$. That however, is impossible since $p_sq_s \geq p_sq_t$, and we would have had by construction $q_s\,H_0^1\,q_t$ and $q_s\,H_0^2\,q_t$, which cannot be reconciled with $q_t\,H_0^1\,q_s$. Thus $(t,s) \in V$, and in fact, since $q_t\,H_0^1\,q_s$, $(t,s) \in V_1$. Following the construction of G, we have an arc from (t,s) to (s,u) and an arc from (s,u) to (t,s) (because $p_sq_s \geq p_s(q_t+q_u)$, $p_tq_t \geq p_tq_s$, $p_uq_u \geq p_uq_s$). Therefore, there is an edge between the vertices (t,s) and (s,u) in G_2, and we conclude that $(s,u) \in V_2$, which implies that $q_s\,H_0^2\,q_u$. This completes the verification of condition (iii).

($\Rightarrow$) Now, we suppose that S is a Yes instance of 2-WARP; there exist H_0^1 and H_0^2 satisfying conditions (i) to (iii). We want to show that the 2-undirected graph G_2 is bipartite. In other words, we want to partition V into two subsets V_1 and V_2 such that there is no edge between two vertices of V_1 and no edge between two vertices of V_2.

Given H_0^1 and H_0^2 we set the vertices in V_1 (respectively in V_2) as follows: a vertex $(s,t) \in V$ belongs to V_1 (respectively to V_2) if $q_s\,H_0^1\,q_t$ (respectively $q_s\,H_0^2\,q_t$). It is not difficult to see that $V_1 \cap V_2 = \emptyset$ and that any vertex in V is either in V_1 or in V_2. Hence, V_1 and V_2 constitute a valid partition of V. We argue, by contradiction, that V_1 and V_2 induce independent sets. Without loss of generality, suppose V_1 is not an independent set. There exist two vertices (s,t) and (u,v) in V_1 with an edge between them in G_2. Thus, in the graph G there is an arc from (s,t) to (u,v), and from (u,v) to (s,t). If both arcs originate from Step 1, we have $u=t$ and $v=s$, which implies $(s,t) \in V_1$, and $(t,s) \in V_1$ which can only happen if $q_s\,H_0^1\,q_t$ and $q_t\,H_0^1\,q_s$; this, however, contradicts condition (ii) for H_0^1. If both arcs originate from Step 2, we also have $u=t$ and $v=s$, and the same argument applies. Hence, one arc originates from Step 1 and one arc originates from Step 2. Without loss of generality, we can assume that the arc from (s,t) to (u,v) comes from Step 1, while the arc from (u,v) to (s,t) comes from Step 2. This implies that $u=t$, and apparently $p_tq_t \geq p_t(q_s+q_v)$. Since $q_s\,H_0^1\,q_t$, condition (iii) implies that $q_t\,H_0^2\,q_v$. By hypothesis, we have $q_t\,H_0^1\,q_v$ and $p_vq_v \geq p_vq_t$ (because $(t,v) \in V_1$). From condition (i) we know that $q_v\,H_0^1\,q_t$ or $q_v\,H_0^2\,q_t$. This, together with $q_t\,H_0^1\,q_v$ and $q_t\,H_0^2\,q_v$, implies that either $q_t\,H_0^1\,q_v$ and $q_v\,H_0^1\,q_t$ or $q_t\,H_0^2\,q_v$ and $q_v\,H_0^2\,q_t$. In the first case, H_0^1 violates condition (ii) whereas in the second case H_0^2 violates condition (ii). In both cases, we have a contradiction with condition (ii). This concludes the proof of Theorem 5. □

5.2. *Algorithm for 2-WARP*

We present an algorithm for 2-WARP that is based on Theorem 5. The pseudocode is described by Algorithm 1.

Algorithm 1 *Algorithm for* 2-WARP

```
1: build the directed graph G from the dataset S
2: build the 2-undirected graph G_2 associated with G
3: if G_2 is bipartite then return Yes, else return No
```

It is clear that each of the three steps of Algorithm 1 can be done in polynomial time. Thus, we have the following result:

Theorem 6. *Algorithm* 1 *solves* 2-WARP *in polynomial time.*

6. Conclusion

We studied three alternative extensions of the weak axiom of revealed preference (WARP) that apply to the collective consumption model. We proved that for the private and public settings, the corresponding testing problem is NP-complete even for two (but also for more) household members. However, for the general setting, testing 2-WARP can be done in polynomial time for households consisting of two members. When there are three or more household members, the complexity of the testing problem for this general setting remains an open question.

Acknowledgments

Laurens Cherchye gratefully acknowledges the European Research Council (ERC) for his Consolidator Grant 614221 and the

Research Fund KU Leuven for the grant STRT1/08/004. Bram De Rock gratefully acknowledges the European Research Council (ERC) for his Starting Grant 263707. Frits Spieksma's research has been partially funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office and FWO Grant G.0447.10.

References

Afriat, S.N., 1967. The construction of utility functions from expenditure data. Internat. Econom. Rev. 8 (1), 67–77.
Apesteguia, Jose, Ballester, Miguel A., 2010. The computational complexity of rationalizing behavior. J. Math. Econom. 46 (3), 356–363.
Cherchye, L., De Rock, B., Platino, V., 2013. Private versus public consumption within groups: testing the nature of goods from aggregate data. Econom. Theory 54, 1248–1258.
Cherchye, L., De Rock, B., Vermeulen, F., 2007. The collective model of household consumption: a nonparametric characterization. Econometrica 75 (2), 553–574.
Cherchye, L., De Rock, B., Vermeulen, F., 2010. Collective Household Consumption Behavior: Revealed Preference Analysis. In: Foundations and Trends in Econometrics, vol. 4. Boston and Delft.
Cherchye, L., De Rock, B., Vermeulen, F., 2011. The revealed preference approach to collective consumption behaviour: testing and sharing rule recovery. Rev. Econom. Stud. 78 (1), 176–198.
Cherchye, L., De Rock, B., Vermeulen, F., 2012. Collective Household Consumption Behavior: Revealed Preference Analysis. In: Foundations and Trends in Econometrics, vol. 4. Now.
Chiappori, P.A., 1988. Rational household labor supply. Econometrica 56 (1), 63–90.
Deb, R., 2010. An efficient nonparametric test of the collective household model. Technical Report. University of Toronto, http://dx.doi.org/10.2139/ssrn.1107246.
Demuynck, T., 2011. The computational complexity of rationalizing boundedly rational choice behavior. J. Math. Econom. 47 (4), 425.
Echenique, F., Lee, S., Shum, M., 2011. The money pump as a measure of revealed preference violations. J. Polit. Econ. 119 (6), 1201–1223.
Garey, M.R., Johnson, D.S., 1979. Computers and Intractability. Vol. 174. Freeman, San Francisco, CA.
Peters, H., Wakker, P., 1994. WARP does not imply SARP for more than two commodities. J. Econom. Theory 62 (1), 152–160.
Piaw, T.C., Vohra, R.V., 2003. Afriat's theorem and negative cycles. Working Paper.
Samuelson, P.A., 1938. A note on the pure theory of consumer's behaviour. Economica 5 (17), 61–71.
Smeulders, B., Spieksma, F.C.R., Cherchye, L., De Rock, B., 2014. Goodness of fit measures for revealed preference tests: complexity results and algorithms. ACM Trans. Econ. Comput. 2 (1), Article 3.
Talla Nobibon, F., Cherchye, L., Crama, Y., Demuynck, T., De Rock, B., Spieksma, F.C.R., 2012. Algorithms for testing the collective consumption model. Working Paper, KU Leuven, Belgium.
Talla Nobibon, F., Cherchye, L., De Rock, B., Sabbe, J., Spieksma, F.C.R., 2011. Heuristics for deciding collectively rational consumption behavior. Comput. Econ. 38 (2), 173–204.
Talla Nobibon, F., Smeulders, B., Spieksma, F.C.R., 2014. A note on testing axioms of revealed preference. J. Optim. Theory Appl. http://dx.doi.org/10.1007/s10957-014-0657-9.
Talla Nobibon, F., Spieksma, F.C.R., 2010. On the complexity of testing the collective axiom of revealed preference. Math. Social Sci. 60 (2), 123–136.
Varian, H.R., 1982. The nonparametric approach to demand analysis. Econometrica 50 (4), 945–973.
Varian, H.R., 2006. Revealed preference. In: Szenberg, M., Ramrattan, L.B., Gottesman, A.A. (Eds.), Samuelsonian Economics and the Twenty-First Century. Oxford University Press, pp. 99–116.
Vermeulen, F., 2002. Collective household models: principles and main results. J. Econ. Surv. 16 (4), 533–564.