Nonparametric tests of collectively rational consumption
behavior: an integer programming procedure

by

Laurens CHERCHYE
Bram DE ROCK
Jeroen SABBE
Frederic VERMEULEN

Public Economics

DISCUSSION PAPER

Nonparametric Tests of Collectively Rational Consumption Behavior: an Integer Programming Procedure

Laurens Cherchye[*], Bram De Rock[†],
Jeroen Sabbe[‡] and Frederic Vermeulen[§]

December, 2007

Abstract

We present an IP-based nonparametric (revealed preference) testing procedure for rational consumption behavior in terms of general collective models, which include consumption externalities and public consumption. An empirical application to data drawn from the Russia Longitudinal Monitoring Survey (RLMS) demonstrates the practical usefulness of the procedure. Finally, we present extensions of the testing procedure to evaluate the goodness-of-fit of the collective model subject to testing, and to quantify and improve the power of the corresponding collective rationality tests.

Key words: collective consumption model, revealed preferences, nonparametric rationality tests, integer programming (IP).

JEL-classification: D11, D12, C14.

1. Introduction

The collective approach, which Chiappori (1988, 1992) originally presented in the context of household labor supply, has become increasingly popular for modeling household consumption behavior. This approach explicitly recognizes that multi-person households consist of several individuals who have their own rational preferences. These individuals jointly take consumption decisions, which are assumed to result in Pareto-efficient intra-household allocations. This collective model provides a positive answer to the methodological and empirical shortcomings of the traditional unitary model, which assumes that multi-person households act as if they were single decision makers.

Browning and Chiappori (1998) provided a characterization of a general collective consumption model, which allows for public consumption and externalities inside the household; they take the "minimalistic" prior that the empirical analyst does not know which commodities are characterized by public consumption and/or externalities. Their core result for two-person households is that under collectively rational behavior the pseudo-Slutsky matrix can be written as the sum of a symmetric negative semi-definite matrix and a rank one matrix. Browning and Chiappori demonstrated necessity of this condition, while Chiappori and Ekeland (2006) addressed the associated sufficiency questions.

The collective rationality test of Browning and Chiappori is parametric in nature; it requires a (non-verifiable) functional/parametric structure that is imposed on the household allocation process and the individual preferences. Cherchye, De Rock and Vermeulen (2007a) established a nonparametric characterization of the same general collective consumption model. More specifically, by using revealed preference axioms, they derived conditions that allow for testing whether observed household consumption behavior is collectively rational, without imposing any parametric structure on the within-houshold decision process (possibly characterized by public consumption and externalities). As such, they also complemented the literature that focuses on nonparametric characterizations and tests of the unitary model; see, for example, Afriat (1967), Varian (1982) and, more recently, Blundell, Browning and Crawford (2003, 2005).

Cherchye, De Rock and Vermeulen (2005) provided a first application to real-life data of these testable nonparametric collective rationality conditions. They test the general collective consumption model for data drawn from the Russia Longitudinal Monitoring Survey (RLMS). The RLMS is one of the few surveys that enables constructing a detailed panel of household consumption, which permits to conduct nonparametric tests without having to assume that preferences are homogeneous across similar individuals in different households. Moreover, there is enough regional and intertemporal relative price variation to test behavioral models in a meaningful way, even though the data contains only 8 observations per household.

While Cherchye, De Rock and Vermeulen (2005) explicitly focused on testing alternative behavioral models (including the unitary model), the current study concentrates

on computational aspects associated with the nonparametric necessity test for collective rationality. Our focus on the necessary condition falls in line with the very nature of the nonparametric approach that we follow, which typically concentrates on the minimal (or 'necessity') empirical restrictions that can be obtained from the available data. See Cherchye, De Rock and Vermeulen (2007a,b) for discussion on complementary sufficiency conditions; our following discussion of the necessary condition readily translates towards these (computationally less complex) alternative conditions.

More specifically, we concentrate on the formulation of the necessity test as an integer programming (IP) problem (see also Cherchye, De Rock and Vermeulen, 2007b). In addition, we show that this IP formulation easily allows for incorporating a number of mechanisms that enhance the computational efficiency of the testing exercise. We assess the practical usefulness of the IP-based test and the corresponding efficiency enhancing mechanisms by means of an empirical application that again uses the RLMS data. Still, given our specific purpose (i.e. *not* testing the validity of behavioral models *in se*), we now do assume preference homogeneity across different households. This assumption allows us to focus on sets of observations that are bigger than those originally considered by Cherchye, De Rock and Vermeulen (2005), and thus to assess computational feasibility of the IP-based necessity test for data sets of reasonable size. Finally, we show that the IP formulation provides a useful basis for evaluating the 'goodness-of-fit' of the collective model subject to testing, and to quantify and improve the 'power' of the corresponding collective rationality tests when brought to real-life data.

The rest of the paper is structured as follows. In Section 2, we reiterate the nonparametric (revealed preference) conditions for rational consumption behavior in terms of the unitary model and, subsequently, the general collective model. Section 3 focuses on operational IP-based procedures to test these nonparametric conditions for collectively rational consumption behavior; this also includes the use of efficiency-enhancing testing mechanisms. Section 4 discusses our application to the RLMS data. Section 5 considers extensions that allow for dealing with goodness-of-fit and power considerations in practical applications. Section 6 concludes.

2. Conditions of rational consumption behavior

2.1. Setting the stage: unitary rationality

We first consider the unitary model for rational household consumption behavior, which models the household as if it were a *single* decision maker; each observed household consumption bundle maximizes a single utility function subject to the corresponding household budget constraint. The unitary nonparametric condition for rational household consumption behavior then essentially requires that there exists such a well-behaved (i.e. non-satiated, concave and continuous) utility function that *rationalizes* the observed household consumption in terms of this unitary model. Assume a situation with N goods, and suppose we observe T household consumption bundles $\mathbf{q}_t \in \mathbb{R}_+^N$

and corresponding prices $\mathbf{p}_t \in \mathbb{R}^N_{++}$ $(t = 1, ..., T)$; let $S = \{(\mathbf{p}_t; \mathbf{q}_t); t = 1, ..., T\}$ be the corresponding set of observations. A utility function U provides a *unitary rationalization* of S if for each observed consumption bundle $\mathbf{q}_t$, with corresponding prices $\mathbf{p}_t$, we have

$$
\begin{aligned}
U(\mathbf{q}_t) \quad &\geq \quad U(\mathbf{q}) \\
\text{for all } \mathbf{q} \quad &\in \quad \mathbb{R}^N_+ \text{ with } \mathbf{p}'_t\mathbf{q} \leq \mathbf{p}'_t\mathbf{q}_t.
\end{aligned}
\tag{2.1}
$$

A core result in the nonparametric literature is that such a unitary rationalization of the data is possible if and only if the data satisfy the *Generalized Axiom of Revealed Preference* (*GARP*; Varian, 1982). Essentially, *GARP* exploits the insight that 'utility maximization over the budget set' can equivalently ('dually') be represented as 'cost minimization over the better-than set'; *GARP* testing boils down to checking this last cost minimization condition.

The distinguishing feature of the nonparametric approach lies in the construction of the better-than set, which is nonparametrically reconstructed as a *revealed preferred* set. First of all, this set includes *directly* revealed preferred bundles. The *direct revealed preference* relation is defined as:

$$
\text{if } \mathbf{p}'_s\mathbf{q}_s \geq \mathbf{p}'_s\mathbf{q}_t, \text{ then } \mathbf{q}_s R_0 \mathbf{q}_t \text{ ('}\mathbf{q}_s \text{ is directly revealed preferred to } \mathbf{q}_t\text{').} \tag{2.2}
$$

The intuition is easy. Under the prices ($\mathbf{p}_s$) and the outlay ($\mathbf{p}'_s\mathbf{q}_s$) that correspond to the bundle $\mathbf{q}_s$, the bundle $\mathbf{q}_t$ was equally obtainable but not chosen. Hence, the household directly reveals its preference for the bundle $\mathbf{q}_s$ over the bundle $\mathbf{q}_t$.

Next, the construction of the revealed preferred set additionally exploits transitivity of preferences. Starting from the notion of direct revealed preference, the more general *revealed preference* relation is defined as:

$$
\begin{aligned}
&\text{if } \mathbf{q}_s R_0 \mathbf{q}_u, \ \mathbf{q}_u R_0 \mathbf{q}_v, ..., \mathbf{q}_z R_0 \mathbf{q}_t \text{ for some (possibly empty) sequence } \mathbf{q}_u, \mathbf{q}_v, ..., \mathbf{q}_z, \\
&\text{then } \mathbf{q}_s R \mathbf{q}_t \text{ ('}\mathbf{q}_s \text{ is revealed preferred to } \mathbf{q}_t\text{').}
\end{aligned}
\tag{2.3}
$$

Clearly, $\mathbf{q}_s R_0 \mathbf{q}_t$ implies $\mathbf{q}_s R \mathbf{q}_t$.

Consistency with the unitary *GARP* condition requires that each observed bundle $\mathbf{q}_t$ must be cost minimizing over the set of revealed preferred bundles $\mathbf{q}_s$ (with $\mathbf{q}_s R \mathbf{q}_t$). Formally, we define this *GARP* condition as:

$$
S \text{ satisfies } GARP \text{ if for all } \mathbf{q}_t: \ \mathbf{p}'_t\mathbf{q}_t \leq \mathbf{p}'_t\mathbf{q}_s \text{ whenever } \mathbf{q}_s R \mathbf{q}_t. \tag{2.4}
$$

Varian (1982) suggested a two-stage procedure for testing data consistency with *GARP*. The first step constructs the revealed preference relations for the set S (using Warshall's algorithm). Subsequently, the second step verifies whether each bundle $\mathbf{q}_t$ is effectively cost minimizing over the correspondingly defined revealed preferred set of observations.

2.2. Collective rationality: the general model

In contrast to the unitary model, the collective model explicitly recognizes the multi-person nature of multi-person households. Moreover, the general collective consumption model allows for (*in casu* positive) externalities and public consumption in this intra-household allocation process. In the present context, public consumption of a certain good, which must be distinguished from private consumption, means that consumption of this good by one household member does not affect the supply available for another household member, and no individual can be excluded from consuming it (at least if one wants to maintain the household). Of course, some commodities may be partly publicly consumed (e.g. car use for a family trip) and partly privately consumed (e.g. car use for work). Next, consumption externalities refer to the fact that one household member gets utility from another member's consumption (e.g. the wife enjoys her husband's nice clothes).

To model externalities and public consumption, we consider *personalized quantities* $\widehat{\mathbf{q}}_t = (\mathfrak{q}_t^1, ..., \mathfrak{q}_t^M, \mathfrak{q}_t^h)$; these personalized quantities decompose each (observed) aggregate consumption bundle $\mathbf{q}_t$ into M quantities $\mathfrak{q}_t^m \in \mathbb{R}_+^N$ capturing the private consumption of each household member m $(m = 1, ..., M)$, and quantities $\mathfrak{q}_t^h \in \mathbb{R}_+^N$ representing public consumption. Of course, the different components of $\widehat{\mathbf{q}}_t$ must add up to the aggregate consumption bundle for each observation t:

$$\mathbf{q}_t = \sum_{m=1}^{M} \mathfrak{q}_t^m + \mathfrak{q}_t^h. \tag{2.5}$$

Each member m has a well-behaved (i.e. non-satiated, concave and continuous) utility function U^m that is non-decreasing in these personalized quantities, which effectively accounts for (positive) externalities and public consumption. The collective model then regards the observed household consumption as the Pareto efficient outcome of a bargaining process between the M household members. Similar to before, a combination of M utility functions $U^1, ..., U^M$ provides a *collective rationalization* of S if for each observed consumption bundle $\mathbf{q}_t$, with corresponding prices $\mathbf{p}_t$, there exist feasible personalized quantities $\widehat{\mathbf{q}}_t$ and weights $\mu_t^1, ..., \mu_t^M \in \mathbb{R}_{++}$ such that:

$$\sum_{m=1}^{M} \mu_t^m U^m (\widehat{\mathbf{q}}_t) \geq \sum_{m=1}^{M} \mu_t^m U^m (\mathfrak{z}^1, ..., \mathfrak{z}^M, \mathfrak{z}^h) \tag{2.6}$$

for all $\mathfrak{z}^1, ..., \mathfrak{z}^M, \mathfrak{z}^h \in \mathbb{R}_+^n$ with $\mathbf{p}_t'[\sum_{m=1}^{M} \mathfrak{z}^m + \mathfrak{z}^h)] \leq \mathbf{p}_t' \mathbf{q}_t$.

In this formulation, each Pareto weight μ_j^m can be interpreted as the *bargaining weight* for the household member m; it represents the weight that is given to this member's utility function in the within-household optimization process.

The nonparametric condition for collectively rational household consumption behavior requires that there exists a representation of each household member's pref-

erences (U^m) and the within-household bargaining process (μ_j^m) that rationalizes the observed household consumption behavior in terms of this general collective model. Cherchye, De Rock and Vermeulen (2007a) established testable (necessary and sufficient) nonparametric conditions for such a collective rationalization of the data. In doing so, they adopted the minimalistic prior that the empirical analyst only observes the aggregate bundle $\mathbf{q}_t$ and not its intra-household allocation; such unobservability is often the case in practical applications. As argued before, our focus is on the testable necessary condition. As we will discuss, this necessary condition has a direct interpretation in terms of the Pareto efficiency assumption that underlies the collective consumption model.

2.3. Pareto efficiency and hypothetical preference relations

The starting point of the nonparametric necessary condition is that the *true* member-specific (revealed) preference relations are *not* observed, because only the aggregate household quantities ($\mathbf{q}_t$) and prices ($\mathbf{p}_t$) can be used. Given this, the condition focuses on so-called *hypothetical member-specific preference relations*. These relations essentially represent *feasible* specifications of the true preference relations in terms of a number of collective rationality conditions ('Rules 1 to 5' in the sequel) defined on the observed (aggregate household) quantities and prices; these conditions have a direct interpretation in terms of Pareto efficiency and individual rationality. The nonparametric condition for collectively rational consumption behavior then requires that *there must exist at least one specification of the hypothetical member-specific preference relations that simultaneously meets all these conditions (i.e. Rules 1 to 5).* Example 1 illustrates this condition (including Rules 1 to 5) for the case with 2 household members ($M = 2$).

To sketch the basic intuition of the hypothetical member-specific preference relations, we first focus on Rule 1, which provides the very basis for constructing these relations. This Rule 1 extends the intuition of the 'unitary' directly revealed preference relation R_0 in (2.2) towards the general collective setting with M household members. Specifically, for the general case with M household members, Pareto efficiency implies:

Rule 1 $\left\{ \begin{array}{l} \text{if } \mathbf{p}_s'\mathbf{q}_s \geq \mathbf{p}_s'\mathbf{q}_t, \text{ then for some } m \in \{1, ..., M\} \text{ we must have} \\ \mathbf{q}_s H_0^m \mathbf{q}_t \text{ ('member } m \text{ hypothetically prefers } \mathbf{q}_s \text{ to } \mathbf{q}_t\text{').} \end{array} \right.$

The intuition is as follows. If the bundle $\mathbf{q}_t$ was equally obtainable under the prices ($\mathbf{p}_s$) and the outlay ($\mathbf{p}_s'\mathbf{q}_s$) that correspond to the chosen bundle $\mathbf{q}_s$, then Pareto efficiency requires that at least one household member must prefer the bundle $\mathbf{q}_s$ to the bundle $\mathbf{q}_t$. If we assume that member m prefers $\mathbf{q}_s$ to $\mathbf{q}_t$, then we specify $\mathbf{q}_s H_0^m \mathbf{q}_t$. Summarizing, the inequality $\mathbf{p}_s'\mathbf{q}_s \geq \mathbf{p}_s'\mathbf{q}_t$ requires that we specify $\mathbf{q}_s H_0^m \mathbf{q}_t$ for at least one m.

The nonparametric condition for collectively rational household consumption requires that there exists at least one specification of the hypothetical revealed preference

relations H_0^m $(m = 1, ..., M)$ that meets Rule 1 defined above and, in addition, a number of additional conditions. We next introduce these additional conditions (Rules 2 to 5).

Rule 2 uses that individual preferences are rational and, thus, transitive. Specifically, we have:

$$\textbf{Rule 2} \; \left\{ \begin{array}{c} \text{if } \mathbf{q}_s H_0^m \mathbf{q}_u, \; \mathbf{q}_u H_0^m \mathbf{q}_v, ..., \mathbf{q}_z H_0^m \mathbf{q}_t \text{ for some} \\ \text{(possibly empty) sequence } \mathbf{q}_u, \mathbf{q}_v, \; ... \; , \mathbf{q}_z, \text{ then } \mathbf{q}_s H^m \mathbf{q}_t. \end{array} \right.$$

The correspondence between the relations H^m and H_0^m for the collective model, which are defined at the level of each individual household member m, is directly analogous to that between the relations R_0 and R for the unitary model, which are defined at the level of the aggregate household.

The following Rules 3 to 5 pertain to rationality *across the household members*. These additional rules use "M^*-*term sums*" of quantities $\sum_{k=1}^{M^*} \mathbf{q}_{t_k}$, which define new bundles as *sums* of observed bundles. Given this, Rule 3 is defined as:

$$\textbf{Rule 3} \; \left\{ \begin{array}{c} \text{for } M^* < M \text{ and } \mathbf{M} \subsetneq \{1, ..., M\}: \text{ if } \mathbf{p}_s' \mathbf{q}_s \geq \mathbf{p}_s' \left(\sum_{k=1}^{M^*} \mathbf{q}_{t_k} \right) \text{ and} \\ \text{for all } m \in \mathbf{M} \text{ we have } \mathbf{q}_{t_{k(m)}} H^m \mathbf{q}_s \text{ for some } k(m) \leq M^*, \\ \text{then } \mathbf{q}_s H_0^l \mathbf{q}_{t_k} \text{ for some } l \in \{1, ..., M\} \setminus \mathbf{M} \text{ and } k \leq M^*. \end{array} \right.$$

This condition expresses that, if all members $m \in \mathbf{M}$ prefer some $\mathbf{q}_{t_{k(m)}}$ over $\mathbf{q}_s$, and the (newly defined) bundle $\sum_{k=1}^{M^*} \mathbf{q}_{t_k}$ is not more expensive than $\mathbf{q}_s$, then the choice of $\mathbf{q}_s$ can be rationalized only if another member $l \notin \mathbf{M}$ prefers $\mathbf{q}_s$ over some $\mathbf{q}_{t_k}$. Indeed, if this last condition were not satisfied, then the bundle $\sum_{k=1}^{M^*} \mathbf{q}_{t_k}$ (under the given prices $\mathbf{p}_s$ and outlay $\mathbf{p}_s' \mathbf{q}_s$) would imply a Pareto improvement over the chosen bundle $\mathbf{q}_s$.

Similarly, we have:

$$\textbf{Rule 4} \; \left\{ \begin{array}{c} \text{for } M^* \leq M \text{ and } \mathbf{M} \subsetneq \{1, ..., M\} : \text{ if } \mathbf{p}_s' \mathbf{q}_s \geq \mathbf{p}_s' \left(\sum_{k=1}^{M^*} \mathbf{q}_{t_k} \right) \text{ and} \\ \text{for all } m \in \mathbf{M} \text{ we have } \mathbf{q}_{t_{k(m)}} H^m \mathbf{q}_s \text{ for some } k(m) \leq M^* - 1, \\ \text{then } \mathbf{q}_s H_0^l \mathbf{q}_{t_{M^*}} \text{ for some } l \in \{1, ..., M\} \setminus \mathbf{M}. \end{array} \right.$$

This condition expresses that, if $\mathbf{q}_s$ is more expensive than the (newly defined) bundle $\sum_{k=1}^{M^*} \mathbf{q}_{t_k}$, while each member $m \in \mathbf{M}$ prefers $\mathbf{q}_{t_{k(m)}}$ (with $k(m) \leq M^* - 1$) over $\mathbf{q}_s$, then the only possibility for rationalizing the choice of $\mathbf{q}_s$ is that another member $l \notin \mathbf{M}$ prefers $\mathbf{q}_s$ over the remaining bundle $\mathbf{q}_{t_{M^*}}$. The interpretation in terms of Pareto efficiency is directly similar to the one for Rule 3.

Rules 1 to 4 define restrictions on the relations H_0^m and H^m. For a specification of these relations, the next condition defines the corresponding upper cost bound

condition:

$$
\textbf{Rule 5} \left\{
\begin{array}{c}
\text{for } M^* \leq M: \\
\text{if for all } m \text{ we have } \mathbf{q}_{s_{k(m)}} H^m \mathbf{q}_t \text{ for some } k(m) \leq M^*, \\
\text{then } \mathbf{p}'_t \mathbf{q}_t \leq \sum_{k=1}^{M^*} \mathbf{p}'_t \mathbf{q}_{s_k}.
\end{array}
\right.
$$

This condition complements Rules 3 and 4. It states that, if each member m prefers $\mathbf{q}_{s_{k(m)}}$ $(k(m) \leq M^*)$ over $\mathbf{q}_t$, then the choice of $\mathbf{q}_t$ can be rationalized only if it is not more expensive than the (newly defined) bundle $\sum_{k=1}^{M^*} \mathbf{q}_{s_k}$. Indeed, if this last condition were not met, then for the given prices $(\mathbf{p}_t)$ and outlay $(\mathbf{p}'_t \mathbf{q}_t)$ all members would be better off by buying the bundle $\sum_{k=1}^{M^*} \mathbf{q}_{s_k}$ rather than the chosen bundle $\mathbf{q}_t$, which of course conflicts with collective rationality.

To summarize, Rules 1 to 5 imply a necessary condition for collectively rational household behavior that can be tested on the available aggregate (price and quantity) information. In fact, it can be verified that the condition reduces to the *GARP* condition for unitary rationality if $M = 1$; i.e. the household consists of a single (e.g. dictatorial) decision maker. For the general case with M members, Cherchye, De Rock and Vermeulen (2007a) have shown that the condition is empirically rejectable as soon as there are $M+1$ goods and $M+1$ observations; e.g., in Example 1, which applies to $M = 2$, we reject collective rationality for a situation with 3 goods and 3 observations.

Example 1. *Consider a situation with 3 goods ($N = 3$) and two household members ($M = 2$), with the following three observed price-quantity combinations ($T = 3$):[1]*

$$
\begin{aligned}
\mathbf{q}_1 &= \begin{pmatrix} 8 & 2 & 1 \end{pmatrix}', \mathbf{q}_2 = \begin{pmatrix} 2 & 1 & 8 \end{pmatrix}', \mathbf{q}_3 = \begin{pmatrix} 1 & 8 & 2 \end{pmatrix}'; \\
\mathbf{p}_1 &= \begin{pmatrix} 5 & 2 & 1 \end{pmatrix}', \mathbf{p}_2 = \begin{pmatrix} 2 & 1 & 5 \end{pmatrix}', \mathbf{p}_3 = \begin{pmatrix} 1 & 5 & 2 \end{pmatrix}'.
\end{aligned}
$$

As a preliminary note, we indicate that this data structure implies

$$
\mathbf{p}'_1 \mathbf{q}_1 > \mathbf{p}'_1 (\mathbf{q}_2 + \mathbf{q}_3); \ \mathbf{p}'_2 \mathbf{q}_2 > \mathbf{p}'_2 (\mathbf{q}_1 + \mathbf{q}_3); \ \text{and } \mathbf{p}'_3 \mathbf{q}_3 > \mathbf{p}'_3 (\mathbf{q}_1 + \mathbf{q}_2).
$$

These inequalities will be useful in our following discussion.

Let us then consider the implications of Rule 1. Given that $\mathbf{p}'_s \mathbf{q}_s \geq \mathbf{p}'_s \mathbf{q}_t$ for each combination $s, t \in \{1, 2, 3\}$, this condition implies $\mathbf{q}_s H^1_0 \mathbf{q}_t$ or $\mathbf{q}_s H^2_0 \mathbf{q}_t$ for each s, t. For example, one possible specification of the hypothetical member-specific preference relations H^1_0 and H^2_0 is:

$$
\begin{aligned}
&\mathbf{q}_1 H^1_0 \mathbf{q}_2, \ \mathbf{q}_1 H^1_0 \mathbf{q}_3 \text{ and } \mathbf{q}_2 H^1_0 \mathbf{q}_3; \\
&\mathbf{q}_3 H^2_0 \mathbf{q}_2, \ \mathbf{q}_3 H^2_0 \mathbf{q}_1 \text{ and } \mathbf{q}_2 H^2_0 \mathbf{q}_1.
\end{aligned}
$$

We will next show that this specification does not simultaneously meet Rules 2 to

---

[1] Cherchye, De Rock and Vermeulen (2007a) used the same data structure to illustrate their necessary condition.

5. A similar inconsistency argument holds for any other specification of H_0^1 and H_0^2 that is consistent with Rule 1.

For the given specification of the relations H_0^1 and H_0^2, Rule 2 implies:

$$\mathbf{q}_1 H^1 \mathbf{q}_2, \ \mathbf{q}_1 H^1 \mathbf{q}_3 \text{ and } \mathbf{q}_2 H^1 \mathbf{q}_3;$$
$$\mathbf{q}_3 H^2 \mathbf{q}_1, \ \mathbf{q}_3 H^2 \mathbf{q}_2 \text{ and } \mathbf{q}_2 H^2 \mathbf{q}_1.$$

Let us then consider Rule 3. Given that $M = 2$, we must only regard $M^* = 1$. For $M^* = 1$, we recall that $\mathbf{p}'_s \mathbf{q}_s \geq \mathbf{p}'_s \mathbf{q}_t$ for each combination $s, t \in \{1, 2, 3\}$. As such, because of Rule 3, $\mathbf{q}_t H^1 \mathbf{q}_s$ implies $\mathbf{q}_s H_0^2 \mathbf{q}_t$ and, similarly, $\mathbf{q}_t H^2 \mathbf{q}_s$ implies $\mathbf{q}_s H_0^1 \mathbf{q}_t$. The given specification of the relations H_0^m and H^m indeed satisfies these requirements.

Next, we consider Rule 4. This rule has no implications for $M^* = 1$, so that we can restrict to $M^* = 2$. Again using that $\mathbf{p}'_s \mathbf{q}_s \geq \mathbf{p}'_s (\mathbf{q}_{t_1} + \mathbf{q}_{t_2})$ for each combination $s, t_1, t_2 \in \{1, 2, 3\}$ with $t_1 \neq t_2$, Rule 4 requires that $\mathbf{q}_{t_1} H^1 \mathbf{q}_s$ implies $\mathbf{q}_s H_0^2 \mathbf{q}_{t_2}$ and, similarly, that $\mathbf{q}_{t_1} H^2 \mathbf{q}_s$ implies $\mathbf{q}_s H_0^1 \mathbf{q}_{t_2}$. The given specification of the relations H_0^m and H^m effectively meets these requirements.

To conclude, we regard Rule 5. First, as for $M^* = 1$, $\mathbf{p}'_s \mathbf{q}_s \geq \mathbf{p}'_s \mathbf{q}_t$ excludes $\mathbf{q}_t H^1 \mathbf{q}_s$ and $\mathbf{q}_t H^2 \mathbf{q}_s$ for any combination s, t; this is indeed the case in our specification of the relations H^m.

Second, as for $M^* = 2$, $\mathbf{p}'_s \mathbf{q}_s > \mathbf{p}'_s (\mathbf{q}_{t_1} + \mathbf{q}_{t_2})$ excludes $\mathbf{q}_{t_1} H^1 \mathbf{q}_s$ and $\mathbf{q}_{t_2} H^2 \mathbf{q}_s$ for any combination $s, t_1, t_2 \in \{1, 2, 3\}$. This requirement is not met for the current specification of hypothetical member-specific preference relations: $\mathbf{q}_1 H^1 \mathbf{q}_2$ and $\mathbf{q}_3 H^2 \mathbf{q}_2$ conflicts with $\mathbf{p}'_2 \mathbf{q}_2 > \mathbf{p}'_2 (\mathbf{q}_1 + \mathbf{q}_3)$. Thus, we conclude that the given specification of the relations H_0^m and H^m does not simultaneously satisfies Rules 1 to 5.

As indicated before, it can be verified that this inconsistency result does not depend on the selected specification of the relations H_0^1 and H_0^2 to obtain consistency with Rule 1. We conclude that the given data structure does not meet the necessary condition for a collective rationalization, i.e. there does not exist a specification of the hypothetical member-specific preference relations H_0^m and H^m that simultaneously meets Rules 1 to 5.

3. Tests of rational consumption behavior

In this section, we show that the nonparametric condition for collectively rational consumption behavior can be verified by solving an integer programming (IP) problem, which implies an operational test for collective rationality that applies for the general case with M household members. This IP formulation was introduced in Cherchye, De Rock and Vermeulen (2007b) and is now operationalized and presented as a programmable code. First, we present the "*basic testing procedure*". Next, we posit that the IP formulation is particularly convenient from a practical point of view, because it allows for implementing (and extending) the efficiency enhancing mechanisms that were presented by Cherchye, De Rock and Vermeulen (2005). This obtains an

"efficiency-enhanced testing procedure".

3.1. Basic testing procedure

In its basic form, the testing procedure is essentially a straightforward three-step algorithm. The first step prepares the available price and quantity data; it summarizes the relevant information in the (aggregate price and quantity) data by identifying dominance relationships involving the M^*-*term sums* that are considered in the different collective rationality rules defined above. The second step formulates the IP problem. The decision variables in this problem are the binary variables x_{st}^m, which correspond to the previously defined hypothetical relations H^m (more specifically, $x_{st}^m = 1$ complies with '$\mathbf{q}_s \, H^m \, \mathbf{q}_t$' and $x_{st}^m = 0$ otherwise). Using this, Rules 1-5 are reformulated in IP terms; the resulting Rules 1-IP to 5-IP define the restrictions of the IP problem. The final step consists of checking whether these IP restrictions define an empty feasible region. An empty feasible region means that the observed household behavior is not collectively rationalizable; a non-empty feasible region indicates that there exists a specification of the hypothetical relations H^m that satisfies Rules 1 to 5.

3.1.1. Summarizing the relevant information

To provide the IP formulation of Rules 1 to 5, we first define the variables $d_{M^*}[s; t_1, ..., t_{M^*}] \in \{0, 1\}$ and $d_{M^*}^+[s; t_1, ..., t_{M^*}] \in \{0, 1\}$ ($M^* \leq M$), as follows:

$$d_{M^*}[s; t_1, ..., t_{M^*}] = 1 \text{ if } \mathbf{p}_s' \mathbf{q}_s \geq \sum_{k=1}^{M^*} \mathbf{p}_s' \mathbf{q}_{t_k} \text{ and } 0 \text{ otherwise;}$$

$$d_{M^*}^+[s; t_1, ..., t_{M^*}] = 1 \text{ if } \mathbf{p}_s' \mathbf{q}_s > \sum_{k=1}^{M^*} \mathbf{p}_s' \mathbf{q}_{t_k} \text{ and } 0 \text{ otherwise.} \qquad (3.1)$$

Basically, these variables compare $\mathbf{p}_s' \mathbf{q}_s$ to the M^*-term sum $\sum_{k=1}^{M^*} \mathbf{p}_s' \mathbf{q}_{t_k}$. This complies with Rules 3 to 5; as we have discussed, these rules compare the expenditure for a single bundle ($\mathbf{q}_s$) with, under the same prices ($\mathbf{p}_s$), the expenditures for a newly defined *sum* of observed bundles ($\sum_{k=1}^{M^*} \mathbf{q}_{t_k}$). (In this respect, also observe that Rule 1 pertains to s and t with $d_1[s; t] = 1$ (i.e. $\mathbf{p}_s' \mathbf{q}_s \geq \mathbf{p}_s' \mathbf{q}_t$).) Obviously, $d_{M^*}^+[s; t_1, ..., t_{M^*}] = 1$ implies $d_{M^*}[s; t_1, ..., t_{M^*}] = 1$; we explicitly define $d_{M^*}^+[s; t_1, ..., t_{M^*}]$ in view of Rule 5-IP introduced below.

For different M^*, the IP formulation will focus on combinations of s and $t_1, ..., t_{M^*}$ for which $d_{M^*}[s; t_1, ..., t_{M^*}] = 1$. Indeed, it follows from our discussion of Rules 1 to 5 that only such combinations define relevant empirical restrictions for the collective rationality test based on Rules 1 to 5; i.e. only such combinations can yield an empirical rejection of collectively rational consumption behavior. As a result, the complexity of the resulting IP problem will be directly related to the structure of the data (as captured by the variables $d_{M^*}^{(+)}[s; t_1, ..., t_{M^*}]$).

3.1.2. Formulation of the IP problem

As mentioned above, the decision variables of the IP problem are the binary variables $x_{st}^m \in \{0,1\}$ ($m = 1, ..., M$; $s, t \in \{1, ..., T\}$), which must be interpreted in terms of the hypothetical preference relations that we had before. Given this, we will next reformulate Rules 1 to 5 in (equivalent) IP terms as 'Rules 1-IP to 5-IP', which are defined on the binary variables x_{st}^m. As such, we can nonparametrically verify data consistency with collective rationality by checking non-emptiness of the feasible set of an IP problem: for a given set of observations S, *there exists a specification of the hypothetical relations H_0^m and H^m that simultaneously satisfies Rules 1 to 5 if and only if there exists a specification of the variables x_{st}^m that simultaneously meets Rules 1-IP to 5-IP*. Example 2 illustrates the IP conditions for the situation described in the previous example.

Rule 1 is equivalent to

$$\textbf{Rule 1-IP}: \sum_{m=1}^{M} x_{st}^m \geq d_1[s;t].$$

This constraint implies that, if $d_1[s;t] = 1$, then $x_{st}^m = 1$ for some m (or $\mathbf{q}_s \, H^m \, \mathbf{q}_t$ for some m).

Rule 2 corresponds to

$$\textbf{Rule 2-IP}: x_{su}^m + x_{ut}^m \leq 1 + x_{st}^m.$$

Thus, if $x_{su}^m = x_{ut}^m = 1$ (or $\mathbf{q}_s \, H^m \, \mathbf{q}_u$ and $\mathbf{q}_u \, H^m \, \mathbf{q}_t$) then $x_{st}^m = 1$ (or $\mathbf{q}_s \, H^m \, \mathbf{q}_t$), which effectively imposes transitivity.

Rule 3 complies with

$$\textbf{Rule 3-IP} \begin{cases} \text{for all } M^* < M \text{ and } \mathbf{M} \subsetneq \{1, ..., M\}: \\ \text{for all } k(m), \ m \in \mathbf{M} \text{ such that } k(m) \leq M^*: \\ d_{M^*}[s; t_1, ..., t_{M^*}] + \sum_{m \in \mathbf{M}} x_{t_{k(m)}s}^m \leq |\mathbf{M}| + \sum_{l \in \{1,...,M\} \setminus \mathbf{M}} \sum_{k=1}^{M^*} x_{st_k}^l. \end{cases}$$

For each combination $t_1, ..., t_{M^*}$ and any subset $\mathbf{M} \subsetneq \{1, ..., M\}$, this constraint considers all possible specifications of the variables $x_{t_{k(m)}s}^m$ with $k(m) \leq M^*$ for all $m \in \mathbf{M}$. For each of these specifications, Rule 3-IP imposes that, if for all $m \in \mathbf{M}$ we have $x_{t_{k(m)}s}^m = 1$ (or $\mathbf{q}_{t_{k(m)}} H^m \mathbf{q}_s$), given $d_{M^*}[s; t_1, ..., t_{M^*}] = 1$, then we must have $x_{st_k}^l = 1$ (or $\mathbf{q}_s H^l \mathbf{q}_{t_k}$) for some $l \notin \mathbf{M}$ and some $k \leq M^*$.

Rule 4 is reformulated as

$$\textbf{Rule 4-IP} \begin{cases} \text{for all } M^* \leq M \text{ and } \mathbf{M} \subsetneq \{1, ..., M\}: \\ \text{for all } k(m), \ m \in \mathbf{M} \text{ such that } k(m) \leq M^* - 1: \\ d_{M^*}[s; t_1, ..., t_{M^*}] + \sum_{m \in \mathbf{M}} x_{t_{k(m)}s}^m \leq |\mathbf{M}| + \sum_{l \in \{1,...,M\} \setminus \mathbf{M}} x_{st_{M^*}}^l. \end{cases}$$

Similar to before, for each combination $t_1, ..., t_{M^*}$ and any subset $\mathbf{M} \subsetneq \{1, ..., M\}$, this constraint considers all possible specifications of the variables $x_{t_{k(m)}s}^m$ with (*in casu*)

$k(m) \leq M^* - 1$ for all $m \in \mathbf{M}$. Rule 4-IP imposes that, if $d_{M^*}[s; t_1, ..., t_{M^*}] = 1$ and for all $m \in \mathbf{M}$ we have $x^m_{t_{k(m)}s} = 1$ (or $\mathbf{q}_{t_{k(m)}} H^m \mathbf{q}_s$), then for the remaining M^* we must have $x^l_{st_{M^*}} = 1$ (or $\mathbf{q}_s H^l \mathbf{q}_{t_{M^*}}$) for some $l \notin \mathbf{M}$.

Finally, Rule 5 corresponds to

$$
\textbf{Rule 5-IP} \quad \left\{ \begin{array}{c} \text{for all } M^* \leq M : \\ \text{for all } k(m) \leq M^* : \\ \sum_{m=1}^{M} x^m_{s_{k(m)}t} \leq M - d^+_{M^*}[t; s_1, ..., s_{M^*}]. \end{array} \right.
$$

For each M^*, this constraint considers all possible specifications of the variables $x^m_{s_{k(m)}t}$ with $k(m) \leq M^*$ for all $m \in \mathbf{M}$; note that, as before, we can have $k(m) = k(l)$ ($m \neq l$) so that $M^* \leq M$. For every possible specification, it requires that, if for each m we have $x^m_{s_{k(m)}t} = 1$ (or $\mathbf{q}_{s_{k(m)}} H^l \mathbf{q}_t$), then it must be that $d^+_{M^*}[t; s_1, ..., s_{M^*}] = 0$ (or $\mathbf{p}'_t \mathbf{q}_t \leq \sum_{k=1}^{M^*} \mathbf{p}'_t \mathbf{q}_{s_k}$).

We conclude that consistency of the observed set S with the general collective rationality condition requires that Rules 1-IP to 5-IP characterize a non-empty feasible region (for $x^m_{st} \in \{0, 1\}$). In the case of a non-empty feasible region, every feasible specification of the binary variables x^m_{st} corresponds to a specification of the relations H^m consistent with Rules 1-5.

Example 2. *We recapture the situation in Example 1, and reformulate Rules 1 to 5 in IP terms. Consistent with our conclusion of Example 1, it can be verified that the correspondingly specified Rules 1-IP to 5-IP define an empty feasible region.*

In this case (with $T = 3$ and $M = 2$) the IP problem has 18 binary variables, namely x^1_{st} and x^2_{st} for each of the 9 possible combinations of $s, t \in \{1, 2, 3\}$. Next, we note that $\mathbf{p}'_s \mathbf{q}_s > \mathbf{p}'_s (\mathbf{q}_{t_1} + \mathbf{q}_{t_2})$ for each combination $s, t_1, t_2 \in \{1, 2, 3\}$. This implies $d_2[s; t_1, t_2] = d^+_2[s; t_1, t_2] = 1$ for all combinations s, t_1, t_2 and, thus, $d_1[s; t] = d^+_1[s; t] = 1$ for all combinations s, t.

Rule 1-IP boils down to

$$\forall s, t : x^1_{st} + x^2_{st} \geq 1.$$

Rule 2-IP is defined as

$$\forall s, t, u : x^1_{su} + x^1_{ut} \leq 1 + x^1_{st} \text{ and } x^2_{su} + x^2_{ut} \leq 1 + x^2_{st}.$$

For Rule 3-IP we only need to consider $M^* = 1$. We define

$$
\begin{aligned}
\text{for } \mathbf{M} &= \{1\} : \forall s, t : 1 + x^1_{ts} \leq 1 + x^2_{st}, \\
\text{for } \mathbf{M} &= \{2\} : \forall s, t : 1 + x^2_{ts} \leq 1 + x^1_{st}.
\end{aligned}
$$

For Rule 4-IP we only need to consider $M^* = 2$, which obtains

$$
\begin{aligned}
\text{for } \mathbf{M} &= \{1\} : \forall s, t_1, t_2 \ (t_1 \neq t_2) : 1 + x^1_{t_1 s} \leq 1 + x^2_{st_2}, \\
\text{and for } \mathbf{M} &= \{2\} : \forall s, t_1, t_2 \ (t_1 \neq t_2) : 1 + x^2_{t_1 s} \leq 1 + x^1_{st_2}.
\end{aligned}
$$

Finally, Rule 5-IP boils down to,

$$\begin{aligned} \text{for } M^* &= 1 : \forall s, t : x_{st}^1 + x_{st}^2 \leq 1, \\ \text{and for } M^* &= 2 : \forall s_1, s_2, t : x_{s_1 t}^1 + x_{s_2 t}^2 \leq 1. \end{aligned}$$

3.1.3. Feasibility check and solution

Once the IP conditions have been generated, the non-emptiness of the feasible region can be checked by using any IP solver. For a non-empty feasible region, every feasible specification of the binary variables x_{st}^m complies with a feasible specification of the hypothetical preference relations H^m.

Strictly speaking, this feasibility check implies a satisfiability problem (also known as a constraint programming problem). The goal is to find at least one feasible solution satisfying rules IP-1 to IP-5. (A more difficult problem is to find all feasible solutions.) In theory, because this is a satisfiability problem, no objective function is needed. However, for some IP solvers an objective function needs to be added in order to perform a feasibility check; this is also the case for Matlab's Bintprog solver that is used in our own empirical application in Section 4. Moreover, adding an objective function can be useful, because it enables us to construct an explicit solution of the hypothetical preference relations H^m that is consistent with Rules 1 to 5.

Of course, multiple specifications of the objective function are usually possible. Because the choice of the objective function is irrelevant for the outcome of the feasibility check, it makes sense to take the easiest possible objective function, i.e. a constant; this is the objective used in our empirical application. Another objective that can be interesting is the following:

$$Min \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{m=1}^{M} x_{st}^m. \tag{3.2}$$

Because $x_{st}^m = 1$ means '$\mathbf{q}_s \ H^m \ \mathbf{q}_t$', this effectively computes the minimal number of hypothetical preference relations that are in the data; this can be interpreted as an indicator of the 'richness' of the data in terms of the preference structure (at the level of the individual household members) that they include. Evidently, a multitude of alternative specifications in terms of the decision variables x_{st}^m are equally possible.

3.1.4. Number of IP restrictions

Essentially, the complexity of the IP problem depends on the number of combinations of M^*, s and $t_1, ..., t_{M^*}$ with $d_{M^*}^{(+)}[s; t_1, ..., t_{M^*}] = 1$. As indicated before, only such combinations can yield an empirical rejection of collectively rational consumption behavior and, therefore, only the corresponding IP restrictions need to be included in the IP problem. As such, we can write the number of IP restrictions as a function of the number of such combinations. This is done in Appendix 1; the functions in that appendix clearly show that the number of relevant restrictions rises exponentially in

T and M.

In this respect, it is worth noting that the basic testing procedure is already more efficient than the "naive" procedure proposed by Cherchye, De Rock and Vermeulen (2007a), which *explicitly* enumerates all possible specifications of the hypothetical member-specific preference relations H_0^m and H^m. As compared to this naive procedure, the use of an IP formulation for the collective rationality test implies *implicit* enumeration of possible specifications of the binary variables x_{st}^m; such implicit enumeration is intrinsic to the Branch-and-Bound procedures that are used by IP solvers (including Matlab's Bintprog solver). By construction, this implicit enumeration exploits the structure of the data more efficiently.

To conclude, we provide some results for (artificial) problems with the largest possible numbers of IP restrictions. Specifically, we consider $M = 2$ (like in our own application), and we specify $d_{M^*}^+[s; t_1, t_2] = 1$ for all s, t_1 and t_s; this necessarily implies a rejection of collectively rationality (e.g. we can reconstruct the argument of Example 1 for every subset of 3 observations). Using the computer configuration of our own empirical application, we obtain a rejection of collective rationality in a CPU-time of 6.87 seconds for $T = 10$, a CPU-time of 73.59 seconds for $T = 15$, and a CPU-time of 432.33 seconds for $T = 20$. (We use $T = 10, 15, 20$ as reference values because our empirical application addresses IP problems based on (subsets with) similar numbers of observations.) This indeed suggests that implicit enumeration efficiently uses the structure of the data.

3.2. Efficiency-enhanced testing procedure

Cherchye, De Rock and Vermeulen (2005) presented two efficiency-enhancing mechanisms that are tailored to the problem at hand. A first mechanism is "*observation filtering*". Essentially, this boils down to filtering out observations that are irrelevant to the collective rationality test; i.e. excluding these observations does not interfere with the test result. This produces a reduced set of observations. A closely related efficiency-enhancing mechanism is "*subsetting*". In essence, this amounts to constructing mutually independent subsets of observations for which the collective rationality test can be carried out separately. In the following, we enrich the basic testing procedure by integrating these two mechanisms. In addition, we propose an extra efficiency-enhancing mechanism that is specific for our IP formulation, which we label "*relation filtering*". Table 3.1 summarizes the resulting efficiency-enhanced testing procedure, which thus adds two 'efficiency-enhancing' steps to the three 'basic' steps discussed before.

3.2.1. Observation filtering and subsetting

The basic idea underlying the *observation filtering* mechanism is that tests for collective rationality need only consider observations k that are implicated in a sequence of

Table 3.1: Efficiency-enhanced testing procedure

Step 1. Summarize relevant information	
Step 1.1. Define all $d_{M^*}[s; t_1, ..., t_{M^*}]$	(basic)
Step 1.2. Observation filtering and subsetting	(efficiency-enhancing)
for each (sub)set of observations:	
Step 2. Formulation of the IP problem	
Step 2.1. Construct IP condition set	(basic)
Step 2.2. Relation filtering	(efficiency-enhancing)
Step 3. Feasibility check and solution	
Step 3.1. Check feasibility of IP condition set	(basic)
Step 3.2. If feasible, give possible solution	(basic)
end for	

observations $Seq(s, t)$ entailing a violation of the unitary $GARP$ condition. The construction of each sequence $Seq(s, t)$ proceeds in two steps. First, for the observations s and t it verifies the condition:

$$\mathbf{q}_s \ R \ \mathbf{q}_t \text{ while } \mathbf{p}'_t \mathbf{q}_t > \mathbf{p}'_t \mathbf{q}_s, \tag{3.3}$$

which implies a violation of the $GARP$ condition for unitary rationality. Next, if condition (3.3) is met, then the corresponding $Seq(s, t)$ includes all observations k with $\mathbf{q}_s \ R \ \mathbf{q}_k$ and $\mathbf{q}_k \ R \ \mathbf{q}_t$. This obtains:

$$
\begin{aligned}
Seq(s, t) &= \{(\mathbf{q}_k; \mathbf{p}_k) \mid \mathbf{q}_s R \ \mathbf{q}_k \text{ and } \mathbf{q}_k R \ \mathbf{q}_t\} \text{ if condition (3.3) applies;} \\
Seq(s, t) &= \emptyset \text{ otherwise.}
\end{aligned}
\tag{3.4}
$$

Observation filtering means that the collective rationality test only considers observations that are in the union *USeq* of the sets $Seq(s, t)$ defined over all couples s and t. Cherchye, De Rock and Vermeulen (2005) argue that (other) observations $(\mathbf{q}'_k; \mathbf{p}'_k) \in S \setminus Useq$ are 'irrelevant' to the test in that they can be omitted without changing the test result.

Subsequently, the *subsetting* mechanism partitions the resulting set *USeq* into subsets that are mutually independent. In this context, mutual independence means that any two subsets $USeq^1$ and $USeq^2$ have no observations in common (i.e. $USeq^1 \cap USeq^2 = \emptyset$). Cherchye, De Rock and Vermeulen (2005) argue that testing the collective rationality condition for each subset separately is equivalent to testing the condition at the level of their union *USeq*.

Appendix 2 presents an algorithm for efficiently implementing observation filtering and subsetting. Because the complexity of the testing problem rises exponentially with the amount of observations, the use of these efficiency-enhancing mechanisms can generate considerable efficiency gains in practice.

3.2.2. Relation filtering

The *relation filtering* mechanism is specific to the IP formulation. It exploits the possibility that the IP problem contains variables x_{st}^m that have a coefficient of zero in all IP restrictions; in terms of Rules 1 to 5 regarding feasible specifications of the hypothetical member-specific preference relations, this means that the aggregate price and quantity data do not imply any restriction on $\mathbf{q}_s H^m \mathbf{q}_t$ for the corresponding member m and the observations s and t. Relation filtering excludes such variables x_{st}^m from the IP test. Such exclusion is efficient because these variable have no relevance for the outcome of the IP test and including them only puts extra burden to the test.

At this point, it is worth noting that, as is easily verified, this relation filtering becomes redundant when observation filtering and subsetting have been carried out. In such a case, the IP problem defined for (testing collective rationality of) each separate subset of observations is such that every variable x_{st}^m has a coefficient equal to one in at least one IP restriction. However, in some extensions of the model (see our discussion in Section 5), observation filtering and subsetting may no longer be appropriate. In those cases, relation filtering can effectively turn out to be a valuable efficiency-enhancing mechanism.

4. Application

As stated in the introduction, a main purpose of the current study is to explore the computational aspects of the IP-based testing procedure presented above. More specifically, we want to demonstrate the practical usefulness of our IP-based test procedure, which also includes illustrating the different efficiency-enhancing testing mechanisms. Ideally, the test should be applied to panel data, where each household is taken as a time series on its own right. This allows for maximal heterogeneity across households and only requires an assumption about homogeneity of preferences over time for a particular household. Cherchye, De Rock and Vermeulen (2005) followed exactly this approach when they conducted nonparametric tests for collective rationality on the Russia Longitudinal Monitoring Survey (RLMS), which is a panel with detailed household consumption. Although this obtained no more than 8 observations per household, the substantial relative price variation across time enabled them to test unitary and collective rationality in a meaningful way.

The aim of the current paper, however, is not a careful analysis of which model fits the data best. Given that we want to illustrate the usefulness of the testing algorithms proposed above, we need more observations per household. As far as we know, existing panel data with detailed consumption only contain a rather limited number of observations per household. For example, Christensen (2007) and Blow, Browning and Crawford (2007) use, respectively, Spanish and Danish consumer panels with at most 24 observations per household. Because we want to demonstrate that the proposed IP-based procedure can handle data sets that are at least of the same order of magnitude, we will assume preference and allocation process homogeneity across

different households; i.e., we will construct data sets that contain quantity and price information of different households that are observed over time. The rest of this section provides a more detailed discussion of the data used in our tests and, subsequently, presents the main results of our empirical analysis.

4.1. Data

Our data are drawn from Phase II of the RLMS, which covers the time period between 1994 and 2003 (Rounds V-XII). The data set contains detailed expenditures and other characteristics from a nationally representative sample of Russian households. Although the RLMS survey design focuses on a longitudinal study of populations of dwelling units, it allows a panel analysis of those households remaining in the original dwelling unit over time.

To keep the analysis simple, the sample selection is for couples with no one else in the household. (We note, however, that the IP test can be applied to the general case with M household members.) We select households where both members are employed, which mitigates the issue of non-separability between consumption and leisure (see Browning and Meghir, 1991). Finally, in order to fully exploit the relative price variation, we only consider households that were observed in all the available rounds of Phase II of the RLMS. This results in a basic sample of 148 couples that are observed 8 times. (More information, including summary statistics, for this basic sample can be found in Cherchye, De Rock and Vermeulen (2005).)

In the empirical application, we focus on a rather detailed commodity bundle that consists of 21 nondurable goods: (1) bread, (2) potatoes, (3) vegetables, (4) fruit, (5) meat, (6) dairy products, (7) fat, (8) sugar, (9) eggs, (10) fish, (11) other food items, (12) alcohol, (13) tobacco, (14) food outside the home, (15) clothing, (16) car fuel, (17) wood fuel, (18) gas fuel, (19) luxury goods, (20) services and (21) rent. Prices are obtained by averaging recorded prices across the households in a given census region. Some of the commodities that we use are aggregate commodities. The price index for a composite commodity is the weighted geometric mean of the prices of the different items in the aggregate good, with weights equal to the average budget shares in a given census region (i.e. the Stone price index).

On the basis of the above sample of 148 households, we construct 200 synthetic data sets that consist of 40 observations each. Every synthetic data set is obtained by randomly drawing 5 households from the basic sample. Since each household is 8 times observed, this results in 40 observations per set. The testing algorithms will be applied to each of these 200 sets. Note that the number of observations per set is larger than for existing consumer panels. Bigger data sets are in principle possible to handle, but there are clearly physical limits (defined in terms of computer memory and speed) on the maximum number of observations. As our following results will indicate, the complexity of the data (defined on the basis of the M^*-term sums, which in turn depend on aggregate quantities and prices) has an important impact on the CPU-time needed to reach a conclusion.

As for a data set that exceeds the physical limits of the IP procedure (for a given computer configuration), a possible solution consists of repeatedly applying the test to randomly drawn subsamples of the original set of observations. If the subsamples are sufficiently small, then such a procedure is always feasible. In addition, it naturally complies with the necessary nature of the collective rationality condition that is subject to testing. (Interestingly, such a repeated subsampling procedure may also provide insight into which (potential 'outlier') observations cause violations of collective rationality.) We will illustrate this application for some data sets in the following. Still, to keep our discussion focused, we abstract from a detailed analysis in this study.

4.2. Results

We programmed the IP-based procedure in Matlab (version 7.4.0.287), because of the matrix-oriented structure of our problem and Matlab's wide availability. The algorithm is available in Matlab m-file format and as a stand-alone application.[2] Any optimization package can be used to solve the generated IP problems; we used Matlab's built-in Integer Programming solver Bintprog on a single computer node of the *K.U.Leuven VIC HPC Cluster* with a two-way 2.4 gHz processor and 16 gigabytes of RAM memory.

Table 4.1 presents summary statistics for the necessity test applied to 200 synthetic data sets consisting of 40 observations each. Focusing on $M = 2$, all data sets pass the necessary condition for data consistency with the collective consumption model; we do not obtain a rejection of collective rationality.[3] (We return to the power of the nonparametric collective rationality tests in Section 5.) Let us then consider the effect of our different efficiency enhancing mechanisms. First, it is clear from the results that the *observation filtering* mechanism is extremely useful. The average number of relevant observations is about 15, which means that on average more than 25 observations per data set turn out to be irrelevant in that they are not (directly or indirectly) involved in a unitary *GARP* violation. The maximum number of relevant observations is 34 and the minimum number is 0; this minimum refers to data sets that can be rationalized by a unitary model.

Next, the *subsetting* mechanism also proves to be very helpful: on average 3 subsets can be constructed by means of the remaining relevant observations. While the

---

[2]The procedure can be downloaded from http://www.kuleuven-kortrijk.be/~u0052996/. While our following application reports results for $M = 2$, this procedure can be used for the general case with M household members. See Sabbe (2007) for details.

[3]The results in Table 4.1 also show that only two synthetic data sets satisfy the unitary *GARP* condition; these are the only synthetic data sets that can be rationalized by means of the unitary model. This deviates from the results in Cherchye, De Rock and Vermeulen (2005), who found that the observed consumption behavior of 117 out of the 148 couples in the basic data set could be rationalized by the unitary model. Of course, the different results in the current study partly follow from the fact that we now focus on synthetic data sets obtained by merging 5 different households. One possible conclusion is that heterogeneity across households should be carefully taken into account when nonparametrically testing behavioral models.

minimum number of subsets is 0 (i.e. the data sets that are consistent with the unitary model), the maximum number is no less than 7. If we have a closer look at these subsets, then we find that the largest subset (which generally requires most of the computation time) contains on average about 9 observations, which is far below the 40 initial observations per data set. Note, however, that there is quite some dispersion around this average: the largest subsets per synthetic data set range from 0 to 34 observations. Given that the necessity test can be computationally burdensome when applied to large data sets, these are interesting results from a practical point of view: they illustrate the operational feasibility of the necessity test applied to real-life data sets.

This also appears from the CPU-time needed to conduct the necessity tests. For each of the 200 synthetic data sets, we imposed a maximum CPU-time of 1 hour to conduct the efficiency-enhanced IP-based testing procedure. The IP-test came to a conclusion within this time limit in 197 cases (i.e. 98.5% of all data sets that we considered). As for the 3 remaining data sets, we experimented with longer maximum CPU-times, but this did not yield a conclusion either. Therefore, we conducted subsampling in these cases: we repeatedly applied the IP-test to randomly drawn subsamples of size 10 and 15; we considered 1000 iterations for each data set. Each of these subsamples was consistent with the necessary condition; we conclude that we cannot reject the necessary condition for these 3 remaining data sets. (Admittedly, our choice of subsample size and number of iterations is rather *ad hoc*; we consider a more in-depth investigation of subsampling design as an interesting avenue for follow-up research.) As for the full collection of 200 data sets, the median CPU-time is about 15 seconds. Once more, there is substantial variation across the synthetic data sets: the minimum is about 1.5 seconds, while the maximum (*in casu* defined over the 197 data sets for which the IP-test reached a conclusion within 1 hour) equals about 43 minutes. All in all, these results are very reasonable.

To analyze the required CPU-time in more detail, we applied a tobit model to account for the right-censoring of 3 CPU-times. The dependent variable is the logarithm of the CPU-time, while the number of subsets, the number of observations in the largest subset and the fraction of *GARP* violations are the independent variables. We are merely interested in associations between the logarithm of the CPU-time and the other variables, although some may argue that the regression even establishes a causal relationship given the set-up of the testing procedure. Next, we recall that our regression pertains to data sets that do pass the collective rationality test; they do not allow drawing conclusions on data sets for which collective rationality is rejected.

Regression results are given in Table 4.2. The number of subsets and the fraction of *GARP* violations turn out to be significantly associated with the dependent variable. Let us then interpret the estimated coefficients (given the small number of censored observations, we interpret marginal effects in terms of the latent variable). *Ceteris paribus*, one extra subset implies an average increase in CPU-time of about 16 per cent. Next, one more observation in the largest subset implies an average increase in

CPU-time of about 2 per cent, all else equal. Finally, a one percentage point higher fraction of $GARP$ violations implies an average increase in CPU-time of about 28 per cent. Note that the constant partly captures the time needed to define the IP-problem. Of course, the definition of the IP-problem also depends on the complexity of the data captured by some of the other regressors. All in all, the regression results are quite intuitive and may provide some guidance to conduct a preliminary analysis of the data before conducting the necessity test. Such a preliminary analysis may give an indication on the feasibility of the necessity test and the time needed to come to a conclusion.

Table 4.1: Computational results

Number of observations	40
Average number of relevant observations	14.62
Minimum number of relevant observations	0
Maximum number of relevant observations	34
Average number of subsets	2.99
Minimum number of subsets	0
Maximum number of subsets	7
Average number of observations in largest subset	9.01
Lowest maximum number of observations in largest subset	0
Highest maximum number of observations in largest subset	34
Median CPU-time in seconds for $M = 2$	14.74
Minimum CPU-time in seconds for $M = 2$	1.5
Maximum CPU-time in seconds for $M = 2$	2554.72
Fraction of sets that satisfy $GARP$	1.00%
Fraction of sets that satisfy collective necessity condition	100%
Fraction of $GARP$ violations	4.14%
Minimum fraction of $GARP$ violations	0.00%
Maximum fraction of $GARP$ violations	32.82%

Averages and medians calculated on the basis of tests applied to 200 synthetic data sets, which contain each 5 households that are 8 times observed. Fraction of sets that satisfy $GARP$ (collective necessity condition) refers to the proportion of sets that pass the associated criterion (with the collective necessity condition pertaining to $M = 2$). Maximum CPU-time is defined over the 197 data sets for which the IP-test reached a conclusion within 1 hour. Fraction of $GARP$ violations equals the average percentage of observations per data set that do not meet the $GARP$ consistency condition (3.3).

Table 4.2: Tobit estimates

Dependent variable: logarithm of CPU-time	Coefficient	Standard error
Constant	1.348*	0.104
Number of subsets	0.155*	0.024
Number of observations in largest subset	0.018	0.017
Fraction of $GARP$ violations	0.281*	0.023
Pseudo R^2	0.67	
Number of observations	200	
Number of right-censored observations	3	
Log likelihood	-123.86	

An asterisk denotes significance at the 1 per cent significance level.

5. Extensions

In the previous section, we have provided an empirical application of our IP-based testing procedure. This demonstrates the practical usefulness of the procedure, even for fairly large data sets. Generally, in view of practical applications two important considerations concern (1) the *goodness-of-fit* of the collective consumption model when data do not pass the ('sharp') condition for collective rationality that is tested, and (2) the *power* of the nonparametric tests in terms of the probability of detecting collectively irrational behavior. For example, in our illustrative application, all data pass the collective rationality tests. This makes the goodness-of-fit concern redundant since the data perfectly fit the (necessary) empirical implications of the collective model under study. Next, as for the power concern, the fact that all data pass the collective rationality test may signal low power of the test.

Conveniently, as we will show, our IP-based procedure allows for incorporating such considerations in the analysis by suitably adapting approaches that have been suggested in the context of nonparametric tests for the unitary rationality model. As for the goodness-of-fit issue, we discuss extensions that evaluate goodness-of-fit in terms of optimization error and in terms of measurement error; this will obtain newly defined programming problems that build on the IP problem formulated in Section 3. Next, when dealing with the power issue, we will also argue that our IP-based approach provides a useful basis for adapting existing approaches to increase the power in practical applications. For compactness, we will not illustrate these extensions in the current study, but the empirical application should be easy given our previous discussion.

5.1. Goodness-of-fit

The collective rationality tests reviewed above are 'sharp' tests; they only tell us whether observations are *exactly* optimizing in terms of the behavioral model that is

under evaluation. However, as argued by Varian (1990), *exact* optimization may not a very interesting hypothesis. Rather, one may be interested whether the behavioral model under study provides a *reasonable* way to describe observed behavior; for most purposes, 'nearly optimizing behavior' is just as good as 'optimizing' behavior. This pleads for using measures that quantify the *goodness-of-fit* of the behavioral model under study. In the following, we discuss two such measures. The first measure is inspired by Varian (1990)'s idea to quantify goodness-of-fit in terms of *optimization error* (which obtains an actual expenditure level that exceeds the -*in casu* collectively-rational level); it can be interpreted as a measure for the *economic significance* of observed violations of collective rationality. The second measure is based on Varian's (1985) 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 collective rationality. To structure our following discussion, we will treat the two measures separately. Still, in practice it can be useful to combine both measures. For example, one may quantify the statistical significance of violations of collective rationality that account for a certain degree of optimization error. Starting from the methodology introduced below, such extensions should be fairly straightforward.

To calculate the goodness-of-fit measures, we endogenously define the variables $d_{M^*}^{(+)}[s; t_1, ..., t_{M^*}] \in \{0, 1\}$ in the programming problem, i.e. we treat them as binary decision variables in our problem formulation. Specifically, for all M^* and corresponding $s, t_1, ..., t_{M^*}$ we include the additional restrictions

$$d_{M^*}[s; t_1, ..., t_{M^*}] \geq \mathbf{p}'_s \left(\widetilde{\mathbf{q}}_s - \sum\nolimits_{k=1}^{M^*} \widetilde{\mathbf{q}}_k \right) + \varepsilon; \qquad (5.1)$$

$$d_{M^*}^+[s; t_1, ..., t_{M^*}] \geq \mathbf{p}'_s \left(\widetilde{\mathbf{q}}_s - \sum\nolimits_{k=1}^{M^*} \widetilde{\mathbf{q}}_k \right).$$

For ε arbitrarily small and positive, this implies $d_{M^*}[s; t_1, ..., t_{M^*}] = 1$ if $\mathbf{p}'_s \widetilde{\mathbf{q}}_s \geq \sum_{k=1}^{M^*} \mathbf{p}'_s \widetilde{\mathbf{q}}_k$; and, analogously, $d_{M^*}^+[s; t_1, ..., t_{M^*}] = 1$ if $\mathbf{p}'_s \widetilde{\mathbf{q}}_s > \sum_{k=1}^{M^*} \mathbf{p}'_s \widetilde{\mathbf{q}}_k$. In this formulation, the vectors $\widetilde{\mathbf{q}}_t \in \mathbb{R}_+^N$ are endogenously defined quantities; they are also treated as decision variables in the programming formulation. Essentially, the following goodness-of-fit measures seek minimal adjustments in the original quantity values, which implies $\widetilde{\mathbf{q}}_t$ that are 'as close as possible' to the observed quantities $\mathbf{q}_t$; the criterion for 'closeness' depends on the specific goodness-of-fit measure at hand.

5.1.1. Optimization error and economic significance

The first measure quantifies *optimization error*; it is inspired on the goodness-of-fit idea of Varian (1990), which is based on Afriat (1972, 1973). Following Varian (1990), this measure quantifies the *economic significance* of observed violations of collective rationality. It seeks the minimal proportional reductions of the observed expenditure levels that is required for establishing consistency with the collective rationality condition. For compactness, our following discussion mainly focuses on the calcula-

tion of such goodness-of-fit measures by starting from the IP formulation discussed in previous sections. We refer to Varian (1990) for a detailed discussion on the interpretation of these measures in practical applications. While Varian focused on the unitary model, his main arguments directly carry over to the general collective model under consideration.

In our formulation, we calculate the reductions in the expenditure levels in terms of proportional reductions of the observed quantities $\mathbf{q}_t$. Specifically, we define for each observation t

$$\widetilde{\mathbf{q}}_t = \theta_t \mathbf{q}_t \text{ with } 0 \leq \theta_t \leq 1; \tag{5.2}$$

again, we treat each variable θ_t as an endogenously defined decision variable. The interpretation is easy: for every observation t, the corresponding value of θ_t captures a proportional expenditure reduction that is independent of the price vector that is used (i.e. $\theta_t = (\mathbf{p}'\widetilde{\mathbf{q}}_t / \mathbf{p}'\mathbf{q}_t)$ for every $\mathbf{p} \in \mathbb{R}_{++}^N$).

Finally, given that we are interested in minimal adjustments of the observed quantity vectors, we can define the objective function of the newly defined programming problem as follows:

$$\max \frac{\sum_{t=1}^{T} \theta_t}{T}.$$

In combination with the decision variables $d_{M^*}^{(+)}[s; t_1, ..., t_{M^*}]$ and $\widetilde{\mathbf{q}}_t$ defined in (5.1) and (5.2), and after adding the Rules IP-1 to IP-5 described above, this obtains a mixed integer linear programming (MILP) problem. This MILP structure implies that the measure can be operationalized, and so provides a useful tool for practical applications to real-life data. (Of course, alternative specifications of the objective function are equally possible; e.g., one can use different weights for the θ_t associated with other observations t.)

The optimal objective function value has a direct interpretation in terms of required expenditure reduction for establishing collective rationality. First, an optimal objective value of unity indicates consistency of observed behavior with the collective rationality condition; in this case, no adjustment of the observed quantities is necessary ($\widetilde{\mathbf{q}}_t = \mathbf{q}_t$ and $\theta_t = 1$ for all t). In the other case, the optimum objective value (below unity) indicates the average expenditure reduction that is required to obtain consistency with the collective rationality conditions; each θ_t gives the corresponding expenditure reduction for every individual observation t. Generally, the objective value can be compared to a specified cut-off level, to assess whether or not observed violations are 'economically significant'; a cut-off level $1 - \alpha$ (e.g. 0.95 or 0.90) then corresponds to a significance level α (e.g. 0.05 or 0.10).

5.1.2. Measurement error and statistical significance

The second measure quantifies *measurement error*; it extends the idea of Varian (1985) to the collective rationality test. This obtains a test for the *statistical significance* of

observed violations of collective rationality. Like before, we will mainly concentrate on the calculation of this goodness-of-fit measure; and we refer to Varian (1985) for a more detailed discussion on its interpretation.

In this case, the vectors $\overline{\mathbf{q}}_t = (\overline{q}_{1;t}, ..., \overline{q}_{N;t})'$ stand for the 'true' quantities, which can be different from the observed quantities $\mathbf{q}_t = (q_{1;t}, ..., q_{N;t})'$. To account for measurement error, we assume the following relationship between true and observed quantities:

$$\overline{q}_{n;t} = q_{n;t} + \eta_{n;t}, \ n = 1, ..., N; \ t = 1, ..., T, \tag{5.3}$$

with the error term $\eta_{n;t}$ assumed to be an independently and identically distributed random variable drawn from $N(0, \sigma^2)$, for σ^2 the variance of the measurement error. Using this, a statistical test for data consistency with the collective rationality model could compute the test statistic

$$\sum_{k=1}^{N} \sum_{t=1}^{T} \frac{(\overline{q}_{n;t} - q_{n;t})^2}{\sigma^2}. \tag{5.4}$$

Under the null hypothesis that the true data satisfy the collective rationality condition, the test statistic follows a Chi-squared distribution with NT degrees of freedom. As such, collective rationality for the data would be rejected if this test statistic exceeded the critical value that corresponds to a specified significance level. However, this test statistic is not observable. Therefore, following Varian (1985), a lower bound on the above statistic can be calculated by means of the programme

$$\min \sum_{k=1}^{N} \sum_{t=1}^{T} \frac{(\widetilde{q}_{n;t} - q_{n;t})^2}{\sigma^2}$$

subject to the vectors $\widetilde{\mathbf{q}}_t = (\widetilde{q}_{1;t}, ..., \widetilde{q}_{N;t})'$ satisfying the necessary condition for collective rationality. Specifically, using the decision variables $d_{M^*}^{(+)}[s; t_1, ..., t_{M^*}]$ in (5.1), and adding the Rules IP-1 to IP-5 obtains a mixed integer quadratic programming (MIQP) problem, which again implies potential operationalization and thus practical usefulness.

Under the null hypothesis, the 'true' data satisfy the constraint, which implies that the resulting function value of the above minimization programme should be no larger than the test statistic (5.4). Consequently, if we reject the null hypothesis on the basis of the obtained function value, then we certainly reject the null hypothesis on the basis of the true test statistic.

In practice, an important difficulty concerns the specification of the variance σ^2. Varian (1985) discusses two alternative solutions. First, we can use estimates of the error variance derived from (parametric or nonparametric) fits of the data, or from knowledge about how accurately the variables were measured. Alternatively, we can

calculate how big the variance needs to be in order the reject to null hypothesis of collectively rational behavior, and compare this to our prior opinions regarding the precision with which the data have been measured.

5.2. Power

In view of empirical applications, an important concern relates to the *power* of the collective rationality tests, i.e. the probability of detecting an alternative hypothesis (*in casu* collectively irrational behavior) to the model under study. Indeed, the fact that the general collective consumption model takes into account externalities and public consumption inside the household, while imposing minimal prior structure, can also make it hardly rejectable in practice. Although our above discussion makes clear that the model can be rejected on the basis of aggregate price and quantity data, the question remains how powerful the theoretical implications are in real-life applications.

Therefore, in addition to the mere nonparametric tests, it generally seems useful to include a power analysis of the tests in practical applications. Bronars (1987) first defined power measures for the unitary rationality model. His alternative hypothesis was based on Becker's (1962) notion of irrational behavior, which states that households randomly choose consumption bundles that exhaust the available budget. Bronars' power measures then capture the probability that the unitary rationality tests reject such randomly drawn consumption bundles from the observed budget hyperplanes. While Bronars focused on the unitary rationality condition, his ideas are readily adapted to the IP-based collective rationality tests presented in the current paper. See also Cherchye, De Rock and Vermeulen (2005) and Cherchye and Vermeulen (2007) for applications of Bronars' approach to nonparametric tests of (less general) collective rationality models. Andreoni and Harbaugh (2006) recently reviewed alternative nonparametric power assessment tools that are currently available, which may also be adapted to the IP test presented here.

If the power turns out to be low, our IP-based procedure provides a useful basis for incorporating alternative strategies to increase that power. First, one can impose additional prior structure on the underlying collective model. In the context of the collective model, this includes using information on observability of the distribution of commodities over household members, or of the publicly consumed quantities; and it also includes the possible use of additional prior structure on the nature of the externalities (e.g. exclude externalities for some or all commodities). A specific application concerns the use of 'assignable commodities', which means that the empirical analyst observes how much a group member consumes (privately) of a particular commodity (e.g. Bourguignon, Browning and Chiappori, 2006); such information on assignable commodities is often used in parametric analysis of collective consumption behavior. See Cherchye, De Rock and Vermeulen (2007b) for the corresponding extensions of the IP formulation used in the current paper; these extensions can be integrated in the IP-based testing procedure that we have presented.

Next, the power of the nonparametric methodology can be further increased by

adapting the 'sequential maximum power path' idea of Blundell, Browning and Crawford (2003, 2005), who focused on unitary rationality. Essentially, the approach of Blundell, Browning and Crawford uses estimated Engel curves for given price regimes to construct 'virtual' quantity bundles that maximize the power of the nonparametric rationality tests. In our opinion, adapting this approach to the methodology presented in this paper can be particularly valuable in view of real-life applications, and thus constitutes an interesting avenue for future research.

6. Conclusion

We have presented an IP-based nonparametric (revealed preference) testing procedure for collectively rational consumption behavior. We focused on the necessary condition derived by Cherchye, De Rock and Vermeulen (2007a) for a general collective consumption model, which accounts for consumption externalities and public consumption while using minimal assumptions on observable price-quantity information. We also showed that the procedure readily allows for incorporating a number of efficiency-enhancing testing mechanisms. Finally, we presented extensions of the testing procedure to evaluate the goodness-of-fit of the general collective consumption model, and to quantify and improve the power of the corresponding collective rationality tests.

An empirical application to households drawn from the Russia Longitudinal Monitoring Survey (RLMS) demonstrated the practical usefulness of the IP-based testing procedure. For realistic numbers of observations (40 observations per synthetic data set in our application), the procedure came to a conclusion in reasonable time. We also found that the efficiency-enhancing mechanisms effectively can (often substantially) reduce the computational burden of the test in practical applications. At this point, it is important to note that the type of tests that we focus on is ideally applied at the level of individual households, to allow for maximal heterogeneity across couples. As stated in the main text, existing detailed consumer panels all have maximum numbers of observations that are substantially below the 40 observations in the synthetic data sets considered in our application, which suggests that the IP-based procedure may also proceed efficiently for other available data.

As for large data sets with IP problems that still exceed the computational limits of the given computer configuration, we recall the possibility of repeatedly applying the test to randomly drawn subsamples of the original set of observations. This procedure is always feasible when the subsamples are sufficiently small; and it complies with the necessary nature of the collective rationality condition that is subject to testing. Still, given the ever increasing computational capacity of computer configurations at a fairly low cost, we may expect that in the near future the procedure will be able to deal with fairly big data sets even without using this subsampling solution.

A final note pertains to the power of the general collective rationality tests that we considered. For example, the fact that our own empirical application did not reject

collective rationality for any data set may signal low power. In turn, this may plead for imposing additional prior structure in practical applications (e.g. in terms of public consumption and externalities within the household). As we have discussed, such extra structure is easily implemented by starting from the IP formulation presented in this paper; see also our reference to the corresponding theoretical specifications of Cherchye, De Rock and Vermeulen (2007b). In this respect, we also recall the possible adaptation of the 'sequential maximum power path' concept of Blundell, Browning and Crawford (2003, 2005) as a possibly fruitful avenue for follow-up research.

Appendix 1

In this appendix, we define the number of IP restrictions as a function of the number of combinations of M^*, s and $t_1, ..., t_{M^*}$ with $d_{M^*}^{(+)}[s; t_1, ..., t_{M^*}] = 1$. For each M^*, let $\#d_{M^*}^{(+)}$ be the number of combinations s and $t_1, ..., t_{M^*}$ with $d_{M^*}^{(+)}[s; t_1, ..., t_{M^*}] = 1$; it follows from our previous discussion that these numbers $\#d_{M^*}^{(+)}$ will define the number of (relevant) IP restrictions associated with Rules IP-1, IP-3, IP-4 to IP-5. Next, let $\#R$ be the number of combinations of s, t and u such that $\mathbf{q}_s R \mathbf{q}_u$ and $\mathbf{q}_u R \mathbf{q}_t$ (in terms of the unitary model: '$\mathbf{q}_s$ is revealed preferred to $\mathbf{q}_u$, and $\mathbf{q}_u$ is revealed preferred to $\mathbf{q}_t$'); this number $\#R$ is relevant for Rule IP-2. Then it can be verified that the different Rules IP-1 to IP-5 imply the following number of (relevant) IP restrictions

for Rule 1-IP: $\#d_1$;

for Rule 2-IP: $\#R \times M$;

for Rule 3-IP: $\sum_{M^*=1}^{M-1} \#d_{M^*} \times \sum_{|\mathbf{M}|=1}^{M-1} [C(M, |\mathbf{M}|) \times (M^*)^{|\mathbf{M}|}]$;

for Rule 4-IP: $\sum_{M^*=1}^{M} \#d_{M^*} \times \sum_{|\mathbf{M}|=1}^{M-1} [C(M, |\mathbf{M}|) \times (M^* - 1)^{|\mathbf{M}|}]$;

for Rule 5-IP: $\sum_{M^*=1}^{M} [(M^*)^M \times \#d_{M^*}^+]$.

We briefly explain how the IP restrictions Rules 1-IP to 5-IP result in these numbers. First, as for Rule 1-IP, we have for each $d_1[s; t] = 1$ exactly one relevant IP restriction. Similarly, for Rule 2-IP, we have for each combination of observations s, t and u that satisfy the condition stated above, M relevant IP restrictions, i.e. one for each member m.

Next, for each $d_{M^*}[s; t_1, ..., t_{M^*}] = 1$, Rule 3-IP implies a relevant IP restriction for each subset $\mathbf{M}$ and each specification of the $k(m)$. The number of subsets $\mathbf{M}$ of given cardinality $|\mathbf{M}|$ is equal to $C(M, |\mathbf{M}|)$, which stands for all drawings without replacement of $|\mathbf{M}|$ elements out of M elements. Given a subset $\mathbf{M}$, the number of specifications of the $k(m)$ is equal to $(M^*)^{|\mathbf{M}|}$, i.e. for each $m \in \mathbf{M}$, $k(m)$ can be any number between 1 and M^*. Hence, to count the number of relevant IP restrictions for a given $d_{M^*}[s; t_1, ..., t_{M^*}] = 1$, we have to multiply these two numbers and consider all possible cardinalities $|\mathbf{M}|$; this obtains the second summation (recall that $\mathbf{M} \subsetneq \{1, ...M\}$). Given that M^* is strictly smaller than M, the first summation then simply considers all possibilities for M^*. Finally, the computation of the numbers of

relevant IP-restrictions stemming from Rules 4-IP and 5-IP is directly similar as for Rule 3-IP.

Appendix 2

In this appendix, we present an algorithm that efficiently implements observation filtering and subsetting. In what follows, *SSM* (for SubSet Matrix) is a matrix which contains all subsets. Each subset is written on a separate line, and empty spaces due to different subset lengths are filled up with zeros. Therefore, the matrix *SSM* has the following dimension: (number of subsets) × (number of observations in the largest subset). The pseudo-code for the algorithm is the following:

$SSM = \emptyset$
for each observation s
    for each observation t
        if $\mathbf{q}_s \ R \ \mathbf{q}_t$ *and* $\mathbf{p}'_t \mathbf{q}_t > \mathbf{p}'_t \mathbf{q}_s$
            $Seq(s,t) = \{k \mid \mathbf{q}_s \ R \ \mathbf{q}_k \text{ and } \mathbf{q}_k \ R \ \mathbf{q}_t\}$;
            if $Seq(s,t) \cap SSM = \emptyset$
                create a new subset: write $Seq(s,t)$ on a new line in SSM;
            else
                define $A = Seq(s,t) \cap SSM$;
                if all $a \in A$ are members of one subset i in SSM
                    add members of $Seq(s,t) \notin A$ to this subset i;
                else *(the elements of A are members of different subsets)*
                    merge those subsets (put them together on one row);
                    add members of $Seq(s,t) \notin A$ to newly created subset in SSM;
                end if...else
            end if...else
        end if
    end for
end for

References

[1] Afriat, S. (1967), "The construction of utility functions from expenditure data", *International Economic Review*, 8, 67-77.

[2] Afriat, S. (1972), Efficiency Estimation of Production Functions, *International Economic Review,* 13, 568-598.

[3] Afriat, S. (1973), "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method", *International Economic Review,* 14, 460-472.

[4] Andreoni, J. and W. Harbaugh (2006), "Power indices for revealed preference tests", *University of Wisconsin-Madison Department of Economics Working Paper* 2005-10, version 2006.

[5] Becker, G. (1962), "Irrational Behavior and Economic Theory", *Journal of Political Economy*, 70, 1-13.

[6] Blow, L., M. Browning and I. Crawford (2007), "Revealed preference analysis of characteristics models", forthcoming in *Review of Economic Studies*.

[7] Blundell, R., M. Browning and I. Crawford (2003), "Nonparametric Engel curves and revealed preferences", *Econometrica*, 71, 205-240.

[8] Blundell, R., M. Browning and I. Crawford (2005), "Best nonparametric bounds on demand responses", *IFS Working Paper*, W05/20.

[9] Bourguignon, F., Browning, M and P.-A. Chiappori (2006), "Efficient intra-household allocations and distribution factors: implications and identification", *CAM Working Paper* 2006-02.

[10] Bronars, S. (1987), "The Power of nonparametric tests of preference maximization", *Econometrica*, 55, 693-698.

[11] Browning, M. and P.-A. Chiappori (1998), "Efficient intra-household allocations: a general characterization and empirical tests", *Econometrica*, 66, 1241-1278.

[12] Browning, M. and C. Meghir (1991), "The effects of male and female labor supply on commodity demands", *Econometrica*, 59, 925-951.

[13] Cherchye, L., B. De Rock and F. Vermeulen (2005), "Opening the black box of intra-household decision-making: theory and non-parametric empirical tests of general collective consumption models", *CentER Discussion Paper* No. 2005-51.

[14] Cherchye, L., B. De Rock and F. Vermeulen (2007a), "The collective model of household consumption: a nonparametric characterization", *Econometrica*, 75, 553-574.

[15] Cherchye, L., B. De Rock and F. Vermeulen (2007b), "The revealed preference approach to collective consumption behavior: testing, recovery and welfare analysis", *CentER Discussion Paper* No. 2007-73.

[16] Cherchye, L. and F. Vermeulen (2006), "Nonparametric analysis of household labour supply: goodness-of-fit and power of the unitary and the collective model", forthcoming in *Review of Economics and Statistics*.

[17] Christensen, M. (2007), "Integrability of demand accounting for unobservable heterogeneity: a test on panel data", *IFS Working Paper*, W14/07.

[18] Chiappori, P.-A. (1988), "Rational household labor supply", *Econometrica*, 56, 63-89.

[19] Chiappori, P.-A. (1992), "Collective labor supply and welfare", *Journal of Political Economy*, 100, 437-467.

[20] Chiappori, P.-A. and I. Ekeland (2006), "The micro economics of group behavior: general characterization", *Journal of Economic Theory*, 130, 1-26.

[21] Sabbe, J. (2007), *COLLIP Manual*, downloadable from http://www.kuleuven-kortrijk.be/~u0052996/.

[22] Varian, H. (1982), "The nonparametric approach to demand analysis", *Econometrica*, 50, 945-972.

[23] Varian, H. (1985), "Nonparametric analysis of optimizing behavior with measurement error", *Journal of Econometrics*, 30, 445–458.

[24] Varian, H. (1990), "Goodness-of-fit in optimizing models", *Journal of Econometrics*, 46, 125–140.