Open Access Repository
www.ssoar.info

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

Cherchye, Laurens; Rock, Bram de; Sabbe, Jeroen; Vermeulen, Frederic

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Cherchye, L., Rock, B. d., Sabbe, J., & Vermeulen, F. (2008). Nonparametric tests of collectively rational consumption behavior: an integer programming procedure. *Journal of Econometrics*, *147*(2), 258-265. https://doi.org/10.1016/j.jeconom.2008.09.010

Nutzungsbedingungen:
Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: http://www.peerproject.eu Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under the "PEER Licence Agreement ". For more Information regarding the PEER-project see: http://www.peerproject.eu This document is solely intended for your personal, non-commercial use.All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-205363

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

Laurens Cherchye, Bram De Rock, Jeroen Sabbe, Frederic Vermeulen

PII: S0304-4076(08)00141-3
DOI: 10.1016/j.jeconom.2008.09.010
Reference: ECONOM 3097

To appear in: *Journal of Econometrics*

Please cite this article as: Cherchye, L., De Rock, B., Sabbe, J., Vermeulen, F., Nonparametric tests of collectively rational consumption behavior: An integer programming procedure. *Journal of Econometrics* (2008), doi:10.1016/j.jeconom.2008.09.010

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

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

April, 2008

Abstract

We present an IP-based nonparametric (revealed preference) testing procedure for rational consumption behavior in terms of a general collective model, which includes 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 (accounting for optimization error as well as measurement error) of the collective model subject to testing.

Key words: collective consumption model, revealed preferences, nonparametric rationality tests, integer programming (IP).
JEL-classification: D11, D12, C14.

---

[*] We are grateful to the special editor Dan Slottje, two anonymous referees and seminar participants in Leuven for helpful comments and suggestions, which substantially improved the paper. We also thank Fabrice Talla Nobibon for helping us with the CPLEX implementation.

[†] University of Leuven, Campus Kortrijk and Center for Economic Studies; Fund for Scientific Research - Flanders (FWO-Vlaanderen). E. Sabbelaan 53, B-8500 Kortrijk, Belgium. E-mail: laurens.cherchye@kuleuven-kortrijk.be.

[‡] European Center for Advanced Research in Economics and Statistics (ECARES), Université Libre de Bruxelles, and ECORE. Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium. E-mail: bderock@ulb.ac.be.

[§] University of Leuven, Campus Kortrijk and Center for Economic Studies; Fund for Scientific Research - Flanders (FWO-Vlaanderen). E. Sabbelaan 53, B-8500 Kortrijk, Belgium. E-mail: jeroen.sabbe@kuleuven-kortrijk.be.

[¶] Corresponding author. CentER, Department of Econometrics and OR, Tilburg University, Netspar and IZA. P.O. Box 90153, NL-5000 LE Tilburg, The Netherlands. E-mail: frederic.vermeulen@uvt.nl. Frederic Vermeulen gratefully acknowledges financial support from the Netherlands Organisation for Scientific Research (NWO) through a VIDI grant.

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. Focusing on a parametric characterization of this general model, they establish that for two-person households collectively rational consumption behavior requires a pseudo-Slutsky matrix that can be written as the sum of a symmetric negative semi-definite matrix and a rank one matrix. Browning and Chiappori show necessity of this condition; Chiappori and Ekeland (2006) address the associated sufficiency question.

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 intra-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 intra-household allocation process and individual preferences (possibly characterized by public consumption and positive 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 on 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. Moreover, there is enough intertemporal relative price variation to test behavioral models in a meaningful way, even though the data contains only 8 observations per household (in casu couples with nobody else in the household).

While Cherchye, De Rock and Vermeulen (2005) explicitly focused on testing alternative behavioral models (including the unitary model), the current study mainly concentrates on operational aspects associated with the nonparametric necessity test for collective rationality. This 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. Generally, such a nonparametric testing analysis provides a valuable first step before imposing more structure to the consumption model under study. In this respect, our discussion for the necessary condition readily extends towards the (complementary) nonparametric sufficiency conditions presented in Cherchye, De Rock and Vermeulen (2007a,b); testing these sufficiency conditions (which have a formally similar structure) is computationally less demanding than testing the necessary condition.

We concentrate on the formulation of the necessity test as a 0-1 Integer Programming (IP) problem, which was proposed by Cherchye, De Rock and Vermeulen (2007b). While the theoretical discussion in Cherchye, De Rock and Vermeulen (2007b) concentrates on the usefulness of this IP formulation for addressing welfare-related questions, we focus on the practical usefulness of the IP-based test for evaluating the 'goodness-of-fit' of the collective model subject to testing. In doing so, we also argue that the IP formulation easily allows for incorporating a number of mechanisms that enhance the computational efficiency of the testing exercise.

Given this specific purpose, we apply the test to the RLMS data discussed above, but now we maintain (or, alternatively, test) the assumption that the intra-household allocation process and individual preferences are homogeneous for couples with males born in the same year. This homogeneity assumption permits 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 the operational feasibility of the IP-based necessity test for data sets of reasonable size. In addition, it demonstrates the usefulness of nonparametric tests for assessing homogeneity assumptions. As a matter of fact, such a test can also be useful from a parametric point of view, given that parametric studies often maintain similar homogeneity assumptions; as such, our empirical application illustrates the value of (complementary) nonparametric collective rationality tests prior to the actual parametric analysis.

At this point, it is worth indicating that our findings for the collective rationality tests can also be insightful in view of designing nonparametric tests that pertain to the unitary model. For example, so far there does not exist a satisfactory (necessary and sufficient) operational test for Varian's (1983) nonparametric weak separability condition. In point of fact, existing unitary separability concepts are formally close to the collective rationality concept under study; see, for example, Blundell and Robin (2000) for a discussion. As such, similar IP-based tests could be conceived for assessing separability in a unitary setting. Our study provides insight into the practical operationalization of such tests.

The rest of the paper is structured as follows. In Section 2, we present the nonparametric (revealed preference) conditions for collectively rational consumption behavior. Section 3 focuses on operational IP-based procedures to test these nonparametric conditions; 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 to evaluate the goodness-of-fit while taking account of optimization error as well as measurement error. Section 6 concludes.

As a final note, we refer to the working paper version (Cherchye et al., 2008) for a detailed description of the presented IP procedure; while the current study focuses on two-person households, the working paper also considers the general setting with M (≥ 2) household members. Sabbe (2007) provides details on the Matlab code that is used in our empirical application.[1]

2 Collectively rational consumption behavior

2.1 The unitary model

To set the stage, we first consider the unitary model for rational household consumption behavior, which models the household as if it were a *single* decision maker. This implies that each observed household quantity bundle is assumed to maximize a sin-

---

[1] Both papers, as well as the Matlab code, can be downloaded from http://www.kuleuven-kortrijk.be/~u0052996/.

gle utility function subject to the corresponding household budget constraint. The unitary nonparametric condition for rational household consumption behavior then essentially requires that there exists a well-behaved (i.e., non-satiated, concave and continuous) utility function that *rationalizes* the observed household consumption in terms of this unitary model.

We assume a situation with N goods and suppose that we observe T household consumption quantity bundles $\mathbf{q}_t \in \mathbb{R}^N_+$ with 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 core result in the nonparametric literature is that a unitary rationalization of the set of observations S is possible if and only if it satisfies the *Generalized Axiom of Revealed Preference* (Varian, 1982).

Definition 1 *Let* $S = \{(\mathbf{p}_t; \mathbf{q}_t); t = 1, ..., T\}$ *be a set of observations. The set* S *satisfies the* Generalized Axiom of Revealed Preference (GARP) *if there exist relations* R_0, R *that meet:*

(i) if $\mathbf{p}'_s\mathbf{q}_s \geq \mathbf{p}'_s\mathbf{q}_t$ *then* $\mathbf{q}_s\ R_0\ \mathbf{q}_t$;

(ii) if $\mathbf{q}_s\ R_0\ \mathbf{q}_u$, $\mathbf{q}_u\ R_0\ \mathbf{q}_v$, ..., $\mathbf{q}_z\ R_0\ \mathbf{q}_t$ *for some (possibly empty) sequence* (u, v, ..., z) *then* $\mathbf{q}_s\ R\ \mathbf{q}_t$;

(iii) if $\mathbf{q}_s\ R\ \mathbf{q}_t$ *then* $\mathbf{p}'_t\mathbf{q}_t \leq \mathbf{p}'_t\mathbf{q}_s$.

In words, the quantities $\mathbf{q}_s$ are 'directly revealed preferred' over the quantities $\mathbf{q}_t$ ($\mathbf{q}_s\ R_0\ \mathbf{q}_t$) if $\mathbf{q}_s$ were chosen when $\mathbf{q}_t$ were equally attainable ($\mathbf{p}'_s\mathbf{q}_s \geq \mathbf{p}'_s\mathbf{q}_t$); see (i). Next, the 'revealed preference' relation R exploits transitivity of preferences; see (ii). Finally, (iii) imposes that the quantities $\mathbf{q}_t$ cannot be more expensive than revealed preferred quantities $\mathbf{q}_s$; otherwise, the household would not act rationally in the unitary sense.

2.2 The general collective consumption model

In contrast to the standard unitary model, the collective model explicitly recognizes the multi-person nature of multi-person households. Moreover, the general collective consumption model under study allows for positive externalities and public consumption in the 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. Of course, some commodities may be partly publicly and partly privately consumed (e.g., car use for a family trip versus 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).

As stated in the introduction, we focus on the case with two household members. Like before, we consider a set of observations $S = \{(\mathbf{p}_t; \mathbf{q}_t); t = 1, ..., T\}$. To model externalities and public consumption, we consider *personalized quantities* $\widehat{\mathbf{q}}_t = (\mathfrak{q}_t^1, \mathfrak{q}_t^2, \mathfrak{q}_t^h)$. These personalized quantities decompose each (observed) aggregate quantity bundle $\mathbf{q}_t$ into quantities $\mathfrak{q}_t^1$ and $\mathfrak{q}_t^2 \in \mathbb{R}_+^N$ capturing the private consumption of each household member 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 quantity bundle for each observation t:

$$\mathbf{q}_t = \mathfrak{q}_t^1 + \mathfrak{q}_t^2 + \mathfrak{q}_t^h.$$

Each member m has a well-behaved 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 two household members. A combination of utility functions U^1 and U^2 provides a *collective rationalization* of S if for each observed quantity bundle $\mathbf{q}_t$, with corresponding prices $\mathbf{p}_t$, there exist feasible personalized quantities $\widehat{\mathbf{q}}_t$ and a weight $\mu_t \in \mathbb{R}_{++}$ such that:

$$\begin{gathered} U^1\left(\widehat{\mathbf{q}}_t\right) + \mu_t U^2\left(\widehat{\mathbf{q}}_t\right) \geq U^1\left(\mathfrak{z}^1, \mathfrak{z}^2, \mathfrak{z}^h\right) + \mu_t U^2\left(\mathfrak{z}^1, \mathfrak{z}^2, \mathfrak{z}^h\right) \\ \text{for all } \mathfrak{z}^1, \mathfrak{z}^2, \mathfrak{z}^h \in \mathbb{R}^n_+ \text{ with } \mathbf{p}_t'[\mathfrak{z}^1 + \mathfrak{z}^2 + \mathfrak{z}^h] \leq \mathbf{p}_t'\mathbf{q}_t. \end{gathered}$$

In this formulation, the weight μ_t can be interpreted as the relative bargaining weight for the second household member; it represents the weight that is given to this member's utility function in the intra-household optimization process.

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 in the introduction, our focus is on the testable necessary condition; we will show that this 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 ($\mathbf{q}_t$) are observed and not the 'true' personalized quantities ($\mathfrak{q}_t^1, \mathfrak{q}_t^2$ and $\mathfrak{q}_t^h$). Given this, the condition focuses on so-called *hypothetical member-*

specific preference relations. These relations essentially represent *feasible* specifications of the true individual preference relations in terms of a number of collective rationality conditions (i.e., conditions *(i)* to *(v)* in Proposition 1 below) defined on the observed (aggregate household) quantities and prices. The nonparametric necessary 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 collective rationality conditions.* The necessary condition is summarized in the following proposition (Proposition 2 of Cherchye, De Rock and Vermeulen, 2007a):

Proposition 1 *Suppose that there exists a pair of utility functions* U^1 *and* U^2 *that provide a* collective rationalization *of the set of observations* $S = \{(\mathbf{p}_t;\mathbf{q}_t)\,;t = 1,...,T\}$. *Then there exist* hypothetical *relations* H_0^m, H^m *for each member* $m \in \{1,2\}$ *such that:*

(i) if $\mathbf{p}'_s\mathbf{q}_s \geq \mathbf{p}'_s\mathbf{q}_t$, *then* $\mathbf{q}_s H_0^1 \mathbf{q}_t$ *or* $\mathbf{q}_s H_0^2 \mathbf{q}_t$*;*

(ii) if $\mathbf{q}_s H_0^m \mathbf{q}_k, \mathbf{q}_k H_0^m \mathbf{q}_l$, ..., $\mathbf{q}_z H_0^m \mathbf{q}_t$ *for some (possibly empty) sequence (k,l,...,z), then* $\mathbf{q}_s H^m \mathbf{q}_t$*;*

(iii) if $\mathbf{p}'_s\mathbf{q}_s \geq \mathbf{p}'_s\mathbf{q}_t$ *and* $\mathbf{q}_t H^m \mathbf{q}_s$, *then* $\mathbf{q}_s H_0^l \mathbf{q}_t$ *(with* $l \neq m$*);*

(iv) if $\mathbf{p}'_s\mathbf{q}_s \geq \mathbf{p}'_s\left(\mathbf{q}_{t_1} + \mathbf{q}_{t_2}\right)$ *and* $\mathbf{q}_{t_1} H^m \mathbf{q}_s$, *then* $\mathbf{q}_s H_0^l \mathbf{q}_{t_2}$ *(with* $l \neq m$*);*

$$(v)\ \begin{cases} a)\ \text{if } \mathbf{q}_s H^1 \mathbf{q}_t \text{ and } \mathbf{q}_s H^2 \mathbf{q}_t, \text{ then } \mathbf{p}'_t\mathbf{q}_t \leq \mathbf{p}'_t\mathbf{q}_s \\ b)\ \text{if } \mathbf{q}_{s_1} H^1 \mathbf{q}_t \text{ and } \mathbf{q}_{s_2} H^2 \mathbf{q}_t, \text{ then } \mathbf{p}'_t\mathbf{q}_t \leq \mathbf{p}'_t(\mathbf{q}_{s_1} + \mathbf{q}_{s_2}) \end{cases} .$$

This condition has a formally similar structure as the unitary *GARP* condition in Definition 1. The essential difference is that Proposition 1 imposes restrictions in terms of 'hypothetical' member-specific preference relations H_0^m and H^m, while *GARP* specifies restrictions in terms 'observable' revealed preference relations R_0 and R.

Condition *(i)* applies to all situations with $\mathbf{p}_s'\mathbf{q}_s \geq \mathbf{p}_s'\mathbf{q}_t$. This means that 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 $\mathbf{q}_s$. In that case, 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. Condition *(ii)* uses that individual preferences are transitive.

The following conditions *(iii)* to *(v)* pertain to rationality *across the household members*. Condition *(iii)* expresses that, if member 1 prefers some $\mathbf{q}_t$ over $\mathbf{q}_s$, and the bundle $\mathbf{q}_t$ is not more expensive than $\mathbf{q}_s$, then the choice of $\mathbf{q}_s$ can be rationalized only if member 2 prefers $\mathbf{q}_s$ over $\mathbf{q}_t$. Indeed, if this last condition were not satisfied, then the bundle $\mathbf{q}_t$ (under the given prices $\mathbf{p}_s$ and outlay $\mathbf{p}_s'\mathbf{q}_s$) would imply an improvement over the chosen bundle $\mathbf{q}_s$.

Similarly, condition *(iv)* states that, if $\mathbf{q}_s$ is more expensive than the (newly defined) bundle $(\mathbf{q}_{t_1} + \mathbf{q}_{t_2})$, while member 1 prefers $\mathbf{q}_{t_1}$ over $\mathbf{q}_s$, then the only possibility for rationalizing the choice of $\mathbf{q}_s$ is that member 2 prefers $\mathbf{q}_s$ over the remaining bundle $\mathbf{q}_{t_2}$. The interpretation in terms of Pareto efficiency is directly similar to the one for condition *(iii)*.

Finally, condition *(v)* complements conditions *(iii)* and *(iv)*; it defines upper cost bounds for each observation t that depend on the specification of the relations H^m. Part *a)* of condition *(v)* states that if both members prefer $\mathbf{q}_s$ over $\mathbf{q}_t$, then the choice of $\mathbf{q}_t$ can be rationalized only if it is not more expensive than $\mathbf{q}_s$. 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 $\mathbf{q}_s$ rather than the chosen bundle $\mathbf{q}_t$, which of course conflicts with Pareto efficiency. Part *b)* of condition *(v)* expresses a similar condition for the case where both members prefer a different bundle $\mathbf{q}_{s_m}$ to $\mathbf{q}_t$. In that case, the choice of $\mathbf{q}_t$ can be rationalized only if it is not more expensive than the bundle $(\mathbf{q}_{s_1} + \mathbf{q}_{s_2})$.

To summarize, conditions *(i)* to *(v)* imply a necessary condition for collectively rational household behavior that can be tested on the available aggregate (price and quantity) information. Cherchye, De Rock and Vermeulen (2007a) show that the condition is rejectable in a two-person setting as soon as there are 3 goods and 3 observations.

3 Nonparametric tests of collective rationality

In this section, we show that the above nonparametric condition for collectively rational consumption behavior can be verified by solving an integer programming (IP) problem. This IP formulation was introduced in Cherchye, De Rock and Vermeulen (2007b); we focus on its practical operationalization. Firstly, we present the *basic testing procedure*. Secondly, we posit that the IP formulation is particularly convenient from a practical point of view, because it allows implementing 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 involves constructing an IP problem and checking whether the feasible region for this problem is empty. The binary decision variables $x_{st}^m \in \{0, 1\}$ of this problem correspond to the previously defined hypothetical relations H^m. For $m = 1, 2$ and $s, t \in \{1, ..., T\}$, we define

$$x_{st}^m = 1 \text{ if } \mathbf{q}_s H^m \mathbf{q}_t \text{ and } x_{st}^m = 0 \text{ otherwise.}$$

Furthermore, we introduce some additional notation that will be used to translate the conditions in Proposition 1 to their IP counterparts:

$$\begin{aligned}
d_1[s;t] &= 1 \text{ if } \mathbf{p}_s'\mathbf{q}_s \geq \mathbf{p}_s'\mathbf{q}_t \text{ and 0 otherwise;} \\
d_1^+[s;t] &= 1 \text{ if } \mathbf{p}_s'\mathbf{q}_s > \mathbf{p}_s'\mathbf{q}_t \text{ and 0 otherwise;} \\
d_2[s;t_1,t_2] &= 1 \text{ if } \mathbf{p}_s'\mathbf{q}_s \geq \mathbf{p}_s'\left(\mathbf{q}_{t_1} + \mathbf{q}_{t_2}\right) \text{ and 0 otherwise;} \\
d_2^+[s;t_1,t_2] &= 1 \text{ if } \mathbf{p}_s'\mathbf{q}_s > \mathbf{p}_s'\left(\mathbf{q}_{t_1} + \mathbf{q}_{t_2}\right) \text{ and 0 otherwise.}
\end{aligned}$$

The IP formulation will solely focus on combinations of observations for which $d_1^{(+)}[s;t]$ $= 1$ or $d_2^{(+)}[s;t_1,t_2] = 1$. Indeed, it follows from our discussion of conditions *(i)* to *(v)* in Proposition 1 that only such combinations define relevant empirical restrictions for the collective rationality test; i.e., only such combinations can yield an empirical rejection of collectively rational consumption behavior.

Given this, it is easy to verify that conditions *(i)* to *(v)* in Proposition 1 can be reformulated in IP terms as conditions *(IP-i)* to *(IP-v)* below. These conditions are defined in terms of the binary variables x_{st}^m.

Proposition 2 *Suppose that there exists a pair of utility functions U^1 and U^2 that provide a* collective rationalization *of the set of observations $S = \{(\mathbf{p}_t;\mathbf{q}_t); t = 1,...,T\}$. Then there exists at least one combination of binary variables $x_{st}^m \in \{0,1\}$ such that for each member $m,l \in \{1,2\}$, we have:*

(IP-i) $\forall s,t : x_{st}^1 + x_{st}^2 \geq d_1[s;t]$;

(IP-ii) $\forall s,t,u : x_{su}^m + x_{ut}^m \leq 1 + x_{st}^m$;

(IP-iii) $\forall s,t : d_1[s;t] + x_{ts}^m \leq 1 + x_{st}^l$ *(with $l \neq m$)*;

(IP-iv) $\forall s,t_1,t_2$ $(t_1 \neq t_2) : d_2[s;t_1,t_2] + x_{t_1 s}^m \leq 1 + x_{st_2}^l$ *(with $l \neq m$)*;

$$\text{(IP-v)}\begin{cases} a)\ \forall s,t: x^1_{st}+x^2_{st}\leq 2-d^+_1[t;s] \\ b)\ \forall s_1,s_2,t: x^1_{s_1t}+x^2_{s_2t}\leq 2-d^+_2[t;s_1,s_2] \end{cases}.$$

Thus, 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^m_0 and H^m that simultaneously satisfies conditions *(i)* to *(v)* in Proposition 1 if and only if there exists a specification of the variables x^m_{st} that simultaneously meets conditions *(IP-i)* to *(IP-v)* in Proposition 2. Strictly speaking, this implies a 0-1 IP feasibility problem; the objective is to find at least one feasible solution satisfying conditions *(IP-i)* to *(IP-v)*.

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: *filtering* and *subsetting*. These mechanisms are easily integrated in the testing procedure. Essentially, these mechanisms reduce the number of observations that are to be considered in the collective rationality test, by exploiting the results of a (computationally easy) unitary *GARP* test preceding the collective rationality test. 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. This will also appear from our own empirical application in Section 4.

The basic idea underlying the *filtering* mechanism is that tests for collective rationality need only consider observations that are implicated in a sequence of observations entailing a violation of the unitary *GARP* condition. A fortiori, only such observations can be involved in a violation of the collective rationality condition. The other observations are irrelevant in that they can be omitted without changing the test result.

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 this context, mutual independence means that any two subsets have no observations in common. 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. Again, this is easily implemented by checking feasibility of a separate IP-problem for each subset. If at least one IP-problem turns out to be infeasible, then collective rationality is rejected.

4 Application

Parametric applications of demand theory typically start from a demand system specification where, in addition to prices and total expenditures, one also controls for demographic variables that influence preferences such as age, schooling level and regional dummies (see, for example, Browning and Meghir, 1991, Banks, Blundell and Lewbel, 1997, and Browning and Chiappori, 1998). The nonparametric counterpart to this approach would be to apply revealed preference conditions to relatively homogeneous subsamples of households (see, for example, Blundell, Browning and Crawford, 2003, and Cherchye and Vermeulen, 2006). With panel data at hand, it is even possible to treat each household as a time series in its own right. This allows for maximal heterogeneity across households and, for a particular household, only requires an assumption about homogeneity of the intra-household allocation process and individual preferences over time. Cherchye, De Rock and Vermeulen (2005) followed 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 couple, the substantial relative price variation across time enabled them to test unitary and collective rationality in a meaningful way.

As stated in the introduction, a main purpose of the current study is to explore the operational aspects of the IP-based testing procedure presented above. More specifically, we want to demonstrate the practical usefulness of our IP-based test procedure. 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 (2008) 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 again make use of the RLMS, but now we maintain (or, alternatively, test) homogeneity of the intra-household allocation process and individual preferences for couples where males share the same birth year. As discussed in the introduction, this also illustrates the usefulness of nonparametric tests for assessing homogeneity assumptions that are frequently used in practice. This can be instrumental for parametric analyses as well. 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.

In the empirical application, we focus on couples with nobody else in the household. Both members are employed in each household that we selected; this mitigates the issue of non-separability between consumption and leisure (see Browning and Meghir, 1991). Next, 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. For each couple, we will focus on a rather detailed consumption 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). For more details on this RLMS data set, including summary statistics, we refer to Cherchye, De Rock and Vermeulen (2005).

On the basis of the aggregate sample of 148 couples, we construct samples of households that contain potentially more than 8 observations. More specifically, we merge all couples of which males share the same birth year. Since we observe 42 different birth years, this implies 42 data sets to which the testing algorithms can be applied separately. As can be seen from Table 1, data set sizes vary from 8 to 128 observations, with on average 28.19 observations per data set; this implies relatively big data sets as compared to the existing consumer panels mentioned above.

[Insert Table 1 around here]

We note that, in principle, the IP procedure can handle any number of observations. But, in practice, for a given computer configuration there will be physical limits (defined in terms of computer memory and speed). As for a data set that exceeds these physical limits, 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. This subsampling procedure will be illustrated in our own empirical application.

4.2 Results

We programmed the construction of our IP problem in Matlab (version 7.4.0.287), because of the matrix-oriented structure of our problem and Matlab's wide availability. Once the IP problem is constructed, any optimization package can be used to solve the problem. We used CPLEX (version 10.2) and the free Matlab interface CPLEXINT to solve the problems on a standard desktop configuration with 1.86 GHz processor and 1 Gb RAM memory.[2]

As a benchmark case, it is interesting to first consider the results for the unitary *GARP* test. We find that only 19 of our 45 data sets (i.e., 45.24%) satisfy *GARP*,

---

[2] See also the Matlab and CPLEX references in our bibliography for more details. As for the IP solver that is used, an obvious choice would have been the Matlab built-in IP solver Bintprog. However, Bintprog performed much worse than CPLEX on our bigger data sets.

which means that they are consistent with the unitary model and, therefore, cannot yield a violation of our collective rationality condition. More than half of the data sets reject the unitary rationality condition; for these data sets our collective rationality condition can be meaningfully tested.

Before turning to these collective rationality tests, it is interesting to assess the effects of the two efficiency enhancing mechanisms that we discussed in Section 3 (which, to recall, exploit the results of the unitary *GARP* test). Table 2 provides a summary of the results; more detailed results are given in the Appendix. First, it is clear from Table 2 that the *filtering* mechanism is extremely useful: the average number of relevant observations (12.79) is far below the average number of observations in the original data sets (28.19); on average, more than 15 observations can be omitted from a data set without changing the result of the collective rationality test. The maximum number of relevant observations is 110 and the minimum number is 0; this minimum refers to data sets that can be rationalized by a unitary model.

[Insert Table 2 about here]

Next, the *subsetting* mechanism also proves to be helpful: Table 2 shows that on average 1.64 subsets can be constructed per data set. While the 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 6. 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 8.71 observations, which is quite below the average of 28.19 initial observations per data set. Note, however, that our results show substantial variation: the number of observations in the largest subset ranges from 0 to 101. 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 show that the efficiency-enhancing mechanisms effectively generate considerable efficiency gains in practice, which of course contributes to the operational feasibility of our IP-based test.

We next turn to the results of the IP-based collective rationality tests. The IP procedure reached a conclusion for all data sets except the largest one with 128 observations, which appeared to be too big for CPLEX to handle due to memory limitations. For each of the other data sets, CPLEX found a solution for the IP problem in less than five minutes of computation time. Once more, there is substantial variation across the data sets: the minimum is less than a second, while the maximum equals almost 5 minutes. All in all, these results are very reasonable, in particular because our IP-based tests were performed on a standard desktop configuration.

For all data sets for which the IP procedure reached a conclusion, the data effectively passed the (necessary) collective rationality condition subject to testing. As for the one data set with 128 observations, we conducted the subsampling procedure suggested above: we repeatedly applied the IP test to randomly drawn subsamples of sizes 50, 60 and 70 (100 replications for each size; subsamples drawn from the largest subset with 101 observations). Each of these subsamples was consistent with the collective rationality condition; and we thus conclude that we cannot reject collective rationality for this remaining data set.

One possible conclusion of these results is that they effectively confirm the assumption of homogeneity of the intra-household allocation process and individual preferences for couples with males born in the same year, which -to recall- is jointly tested with collective rationality in our application. Given our specific selection of couples (with both household members employed, and nobody else in the household), this could indeed be a valid interpretation. An alternative (and complementary) conclusion pertains to the generality of the model that is subject to testing, which implies minimal *a priori* structure on the intra-household allocation process. Such a general model may induce low power (i.e., a low probability of detecting collectively irrational behavior). From this perspective, the IP-based test under study can be considered as a useful first step of a more focused analysis; in such a set-up, subsequent steps can impose additional structure on the collective decision model. We return to these power-related issues in the concluding section.

5 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 be 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 our illustrative application, all data pass the collective rationality tests. This makes the goodness-of-fit concern redundant in this case, since the data perfectly fit the (necessary) empirical implications of the collective model under study.

Still, it is worth indicating that our IP-based testing methodology easily allows for taking such goodness-of-fit concerns into account for data sets that do reject the collective rationality condition. Specifically, we consider two goodness-of-fit measures that have been suggested in the literature on nonparametric tests for the unitary model; we translate these measures towards our collective set-up. The first measure is inspired by Varian's (1990) 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 $\widetilde{d}_1^{(+)}[s;\, t]$ and $\widetilde{d}_2^{(+)}[s;\, t_1, t_2] \in \{0, 1\}$ in the programming problem; i.e., we treat them as binary decision variables in our problem formulation. Specifically, for all s, t, t_1, t_2 we include the additional restrictions

$$\begin{aligned} \widetilde{d}_1[s;t] &\geq \mathbf{p}_s' \left(\widetilde{\mathbf{q}}_s - \widetilde{\mathbf{q}}_t \right) + \varepsilon; \\ \widetilde{d}_1^{+}[s;t] &\geq \mathbf{p}_s' \left(\widetilde{\mathbf{q}}_s - \widetilde{\mathbf{q}}_t \right); \\ \widetilde{d}_2[s;t_1,t_2] &\geq \mathbf{p}_s' \left(\widetilde{\mathbf{q}}_s - \left(\widetilde{\mathbf{q}}_{t_1} + \widetilde{\mathbf{q}}_{t_2} \right) \right) + \varepsilon; \\ \widetilde{d}_2^{+}[s;t_1,t_2] &\geq \mathbf{p}_s' \left(\widetilde{\mathbf{q}}_s - \left(\widetilde{\mathbf{q}}_{t_1} + \widetilde{\mathbf{q}}_{t_2} \right) \right). \end{aligned} \tag{1}$$

For any ε arbitrarily close to zero and positive, this implies $\widetilde{d}_1[s;t] = 1$ if $\mathbf{p}_s'\widetilde{\mathbf{q}}_s \geq \mathbf{p}_s'\widetilde{\mathbf{q}}_t$ and $d_1^+[s;t] = 1$ if $\mathbf{p}_s'\widetilde{\mathbf{q}}_s > \mathbf{p}_s'\widetilde{\mathbf{q}}_t$ (and analogously for $\widetilde{d}_2^{(+)}[s;\, t_1, t_2]$). 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 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). 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 calculation of such goodness-of-fit measures by starting from the IP formulation discussed in the previous section. 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

$$\widetilde{\mathbf{q}}_t = \theta_t \mathbf{q}_t \text{ with } 0 \leq \theta_t \leq 1. \tag{2}$$

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 $\widetilde{d}_1^{(+)}[s;t]$ and $\widetilde{d}_2^{(+)}[s;t_1,t_2]$ and $\widetilde{\mathbf{q}}_t$ defined in (1) and (2), and after adding the conditions *(IP-i)* to *(IP-v)* in Proposition 2, 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.

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).

5.2 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. (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:

$$\overline{q}_{n;t} = q_{n;t} + \eta_{n;t} \text{ for } n = 1, ..., N \text{ and } t = 1, ..., T,$$

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

$$\sum_{n=1}^{N} \sum_{t=1}^{T} \frac{\left(\overline{q}_{n;t} - q_{n;t}\right)^2}{\sigma^2}. \tag{3}$$

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

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

subject to the vectors $\widetilde{\mathbf{q}}_t = \left(\widetilde{q}_{1;t}, ..., \widetilde{q}_{N;t}\right)'$ satisfying the necessary condition for collective rationality. Specifically, using the decision variables $\widetilde{d}_1^{(+)}[s;t]$ and $\widetilde{d}_2^{(+)}[s;t_1,t_2]$ in (1), and adding the conditions *(IP-i)* to *(IP-v)* in Proposition 2 obtains a mixed integer quadratic programming (MIQP) problem, which again implies 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 (3). 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.

6 Concluding discussion

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; when data do not pass the 'sharp' condition for collective rationality, such a goodness-of-fit analysis is easily incorporated in the IP formulation. As discussed in the introduction, our findings for IP-based tests of the collective model can also be useful to conceive IP-based procedures for testing within a unitary setting (e.g., testing specific separability assumptions).

An empirical application to households drawn from the Russia Longitudinal Monitoring Survey (RLMS) demonstrated the practical usefulness of the IP-based testing procedure. Specifically, using a maintained assumption that the intra-household allocation process and individual preferences are homogeneous for couples with males born in the same year, we constructed 42 data sets containing between 8 and 128 observations; we conducted the IP-based test for each data set separately. Firstly, we found that the efficiency-enhancing mechanisms effectively can (often substantially) reduce the computational burden of the test in practical applications. Next, using a standard desktop configuration, our IP-based collective rationality tests came to a conclusion in less than five minutes for all but one of our 42 data sets. For the one remaining data set the IP problem exceeded the computational limits of our desktop configuration; in this case, we performed a subsampling procedure that repeatedly applies 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.

We could not reject collective rationality for any of the 42 data sets. One possible conclusion is that the jointly tested collective rationality and homogeneity assumptions effectively do hold for the data sets under study; given our specific selection criteria, which obtain relatively homogeneous data sets, this may indeed be a valid interpretation. Alternatively, the fact that all data pass the collective rationality tests may signal low power (i.e., low probability of detecting collectively irrational behavior). Indeed, the general collective model imposes minimal prior structure, which can make it hardly rejectable in practice. Although the nonparametric collective rationality condition under study can clearly be rejected on the basis of aggregate price and quantity data, the question remains how powerful the theoretical implications are in real-life applications. Such power considerations are especially relevant when the main focus is on testing specific behavioral hypotheses as such, rather than on operational aspects, as in this study.

As for practical applications in which a power analysis is recommendable, it is worth noting that the presented IP-based collective rationality tests readily include power measures that have been suggested in a unitary framework (e.g., Bronars, 1987, and Andreoni and Harbaugh, 2006); see, e.g., Cherchye, De Rock and Vermeulen (2005) and Cherchye and Vermeulen (2006) for such power assessments of (less general) collective models. Next, if the power turns out to be low, additional prior structure can be imposed 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 the corresponding theoretical specifications in Cherchye, De Rock and Vermeulen (2007b). Finally, the power of the nonparametric collective rationality tests could be further increased by suitably adapting the 'sequential maximum power path' idea of Blundell, Browning and Crawford (2003, 2005), who originally focused on a unitary setting.

Appendix: Details on tested data sets

Birth year	Nr. of obs.	Nr. of relevant obs.	Nr. of subsets	Nr. of obs. per subset
1918	8	3	1	3
1919	8	0	0	0
1920	8	0	0	0
1922	8	0	0	0
1923	16	0	0	0
1924	40	26	4	2;2;7;15
1925	8	0	0	0
1926	48	18	4	2;2;3;11
1927	56	50	5	3;4;4;12;27
1928	24	9	2	2;7
1929	64	43	4	3;3;5;32
1930	64	43	5	2;2;6;16;17
1931	40	4	2	2;2
1932	40	13	3	2;5;6
1933	16	4	1	4
1934	16	0	0	0
1935	128	110	3	2;7;101
1936	80	53	6	2;2;2;2;5;40
1937	56	32	4	2;4;5;21
1938	64	41	6	2;4;4;6;9;16
1939	48	2	1	2
1940	56	25	4	2;2;3;18
1941	48	31	3	2;3;26
1942	16	0	0	0
1943	8	0	0	0
1944	8	0	0	0
1945	24	4	2	2;2
1946	16	0	0	0
1947	16	0	0	0
1948	24	10	3	2;4;4
1949	16	2	1	2
1950	24	5	1	5
1951	8	0	0	0
1953	8	0	0	0
1954	16	4	2	2;2
1955	8	0	0	0
1957	8	0	0	0
1960	8	0	0	0
1962	8	5	2	2;3
1964	8	0	0	0
1969	8	0	0	0
1972	8	0	0	0

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] Banks, J., R. Blundell and A. Lewbel, 1997, Quadratic Engel curves and consumer demand. *Review of Economics and Statistics* 79, 527-539.

[6] Blow, L., M. Browning and I. Crawford, 2008, Revealed preference analysis of characteristics models. *Review of Economic Studies* 75, 371-389.

[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] Blundell, R. and J.-M. Robin, 2000, Latent separability: grouping goods without weak separability. *Econometrica* 68, 53-84.

[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, J. Sabbe and F. Vermeulen, 2008, Nonparametric tests of collectively rational consumption behavior: an integer programming procedure. *CentER Discussion Paper* No. 2008-2.

[14] 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.

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

[16] 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.

[17] 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*.

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

[19] Chiappori, P.-A., 1988, Rational household labor supply. *Econometrica* 56, 63-89.

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

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

[22] CPLEX 10.2, ILOG Inc., 2008, www.cplex.ilog.com.

[23] CPLEXINT, Matlab interface for the CPLEX solver, downloadable from http://control.ee.ethz.ch/hybrid/cplexint.php

[24] Matlab 7.4, R2007a, The MathWorks Inc., http://www.mathworks.com.

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

[26] Varian, H., 1982, The nonparametric approach to demand analysis. *Econometrica* 50, 945-972.

[27] Varian, H., 1983, Non-parametric tests of consumer behaviour. *Review of Economic Studies* 50, 99-110.

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

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

Table 1: Frequency table for data set sizes

Number of observations	Frequency
8	16
16	8
24	4
40	3
48	3
56	3
64	3
80	1
128	1
Total	42

Table 2: Descriptive statistics after use of efficiency-enhancing mechanisms

	Average	Std. dev.	Min.	Max.
Number of observations per data set	28.19	26.22	8	128
Number of relevant observations per data set	12.79	22.09	0	110
Number of subsets per data set	1.64	1.91	0	6
Number of observations in largest subset	8.71	17.72	0	101