

Collective consumption models with restricted bargaining weights: an empirical assessment based on experimental data

May 2011

Abstract

We use experimental data to analyze consumption decisions by groups of individuals who have to reach a consensus on spending a joint budget. Our experiment involves dyads (i.e. two-member groups) who have to compose consumption bundles consisting of three commodities (wine, orange juice and M&Ms). We focus on the collective consumption model to describe group behavior. This model represents group decisions as Pareto optimal outcomes of a within-group bargaining process, with rational group members who are each characterized by individual bargaining weights. We also consider specifications of the collective model that restrict the variation of these bargaining weights. A distinguishing feature of our study is that we use revealed preference testing tools to assess the goodness-of-fit and discriminatory power of alternative specifications of the collective model. Our experimental results suggest that the most appropriate model specification allows for a limited variation of the bargaining weights.

Key words: Pareto efficiency, collective consumption decisions, revealed preferences, experimental data, bargaining weights.

JEL-classification: C14, C92, D11, D12, D13.

1 Introduction

We consider the empirical analysis of consumption decisions by groups of individuals who have to reach a consensus on spending a joint budget. A most notable example of such a group is the household of which the members pool their income,

after which they must agree upon the allocation of this aggregate income. Samuelson (1956) initiated this consensual approach to modeling household consumption behavior, by assuming a (household level) social welfare function that is maximized under a household budget constraint (see also Donni (2007a) and Lundberg and Pollak (2007) for recent surveys of the literature following this approach).¹

This paper deals with individual consumption in a multi-person group and is framed in terms of the collective consumption model introduced by Chiappori (1988). This model assumes a collective choice process and explicitly recognizes the multi-person nature of this process, with each individual decision maker (group member) characterized by an own utility function that represents her/his rational preferences.² It only assumes that the observed group consumption is the Pareto efficient outcome of a bargaining process.³ The model then defines collectively rational behavior as maximizing a weighted sum of the group member utility functions, with the weights representing the bargaining power of the individual group members. Interestingly, these bargaining weights may vary depending on the prices, income levels and other exogenous variables characterizing the choice situations.⁴ The collective model is increasingly used in empirical studies of household behavior (see, for example, Vermeulen (2002) and Donni (2007b) for surveys).

We make an empirical as well as a methodological contribution. The empirical contribution is that we apply a revealed preference methodology in combination with experimental data for assessing the empirical goodness-of-fit of the collective model. More specifically, we will focus on revealed preference conditions in the

¹In this respect, it is worth pointing out that alternative approaches to modeling household behavior have been suggested in the literature. For example, Grossbard-Shechtman (1984, 2003) and Grossbard (2010) assume individuals who independently maximize their own utility functions under individual budget and participation constraints. Still, in our experimental design respondents face (only) a joint budget constraint (see Section 3). Therefore, our principal focus will be on the consensual approach in what follows.

²This explicit recognition of individual preferences falls in line with the growing consensus that multi-person household consumption behavior should no longer be treated as if the household were a single decision maker (optimizing a household utility function subject to a household budget constraint). Indeed, this so-called unitary model of household consumption imposes empirically testable restrictions on the household demand function (e.g. Slutsky symmetry) that are frequently rejected when confronted with consumption or labor supply data (see, for example, Thomas (1990), Browning and Chiappori (1998), Duflo (2003) and Cherchye, De Rock and Vermeulen (2009) for empirical evidence based on consumption data, and Lundberg (1988), Fortin and Lacroix (1997), Chiappori, Fortin and Lacroix (2002) and Cherchye and Vermeulen (2008) for evidence based on labor supply data).

³Pareto efficiency is a minimal assumption in the consensual approach to modeling group consumption behavior. Apps and Rees (1988, 2009) propose a household consumption model, which is formally similar to Chiappori's model, based on the same Pareto efficiency assumption.

⁴These variables are usually referred to as distribution factors (see, for example, Browning, Chiappori and Lechene (2006) for an extensive discussion). They have also been termed market forces (in Becker, 1973) and extra environmental parameters (in McElroy and Horney, 1981).

tradition of Afriat (1967)-Diewert (1973)-Varian (1982). These conditions enable checking consistency of a given data set with a particular specification of the collective consumption model. In the spirit of Varian (1982), we will refer to this checking procedure as ‘testing’ data consistency with collective rationality. Apart from goodness-of-fit, we will also consider the discriminatory power of alternative specifications of the collective consumption model. Indeed, a fair comparison of different behavioral models must complement a goodness-of-fit analysis with a power analysis: favorable goodness-of-fit results, indicating few violations of the behavioral restrictions, have little meaning if the behavioral implications have low power, i.e. optimizing behavior can hardly be rejected.

Our use of revealed preference tests and experimental data distinguishes our empirical study from existing studies. Revealed preference tests are entirely non-parametric, which means that they do not require a prior parametric structure for the consumption model (e.g. individual utility functions). This contrasts with other studies, which are typically parametric in nature. Strictly speaking, these studies simultaneously test the consumption model under study as well as a (non-verifiable) parametric structure that is imposed on the model. The use of nonparametric tools does not require such a priori’s. In addition, the laboratory nature of experiments effectively avoids the often controversial preference homogeneity assumptions (excluding e.g. changing preferences) and data measurement problems that are associated with using ‘real life’ data. In fact, it has been argued that revealed preference testing tools are especially useful within an experimental context; a particularly convincing case is provided by Sippel (1997), who focused on individual rationality. Next, and specific for our own study, the experimental set-up allows for obtaining information on consumption quantities for the individual group members; such information is typically not available in ‘real life’ data sets. For example, household data sets usually only contain consumption quantity information at the level of the aggregate household as a whole, and do not reveal the individual members’ consumption quantities. In this respect, we also refer to our discussion in Section 4, which discusses the relevance of this study for analyzing household consumption data.

As for our methodological contribution, we present testing tools for collective rationality under alternative assumptions regarding variation of bargaining weights across different choice situations. This extends earlier work of Cherchye, De Rock and Vermeulen (2007, 2011), who developed revealed preference conditions for the collective consumption model that do not include such assumptions. As such, these conditions allow for a huge variation of the bargaining weights

across periods. For example, the model allows the full bargaining power to shift from one group member to another between any two consecutive periods. Because the realistic nature of such power shifts may be questioned, we propose a revealed preference methodology for testing the collective consumption model under restricted variation of bargaining weights. As we will indicate, adding such weight restrictions implies revealed preference tests with higher discriminatory power, which is particularly convenient when focusing on the empirical performance of the collective model.

At this point, it is worth to distinguish our approach from the one of Cherchye, De Rock and Vermeulen (2009). These authors suggest revealed preference conditions for the collective consumption model that impose prior assumptions regarding the so-called sharing rule (or within-group income distribution) underlying observed household behavior. This sharing rule is often interpreted as an indicator of bargaining power (see, for example, Browning, Chiappori and Lewbel (2006)). As such, the approach presented here can be conceived as complementary with the one proposed by Cherchye, De Rock and Vermeulen.⁵

One final remark applies to our following empirical analysis. The revealed preference tests on which we focus are not traditional statistical tests, which are characterized by standard errors and so allow for statistical inference. Our tests are ‘sharp’ tests: they check whether or not the data pass the revealed preference conditions exactly. If the data do not pass the conditions, then the model under study is rejected.⁶ As a result, we cannot use the usual statistical methods for comparing the empirical validity of different specifications of the collective model. By contrast, we will follow a recent proposal of Beatty and Crawford (2011) that compares different behavioral models on the basis of a so-called ‘predictive success’ measure, which is specially tailored for the type of revealed preference tests that we consider here. As we will explain in Section 4, this measure of predictive success simultaneously accounts for the goodness-of-fit and discriminatory power of a particular model specification.

The rest of the paper unfolds as follows. Section 2 introduces the revealed preference tests for collective rationality. Section 3 presents the experimental design. Section 4 presents the results of our empirical analysis, and subsequently discusses the usefulness of this study for analyzing household consumption data.

⁵For example, it may well be that the distribution of the bargaining power varies while the within-group income distribution remains constant, and vice versa. Example 1 in Section 2.3 will provide a specific illustration of this point.

⁶Varian (1990) provides a detailed discussion on the difference between revealed preference tests and traditional statistical tests.

Section 5 summarizes and concludes.

2 Rational consumption decisions: revealed preference conditions

In this section we present the rationality conditions for consumption behavior that will be used in our empirical study. To set the stage, we first introduce the unitary rationality condition, which we will also use for evaluating the individual choices in our experimental analysis. Next, we consider two versions of collective rationality, i.e. with and without restrictions on the variation of the bargaining weights across choice observations.

2.1 Unitary rationality

Our empirical analysis starts from a finite set of T observed choices consisting of N -vectors of quantities $\mathbf{q}_t \in \mathbb{R}_+^N$ and prices $\mathbf{p}_t \in \mathbb{R}_{++}^N$. Let

$$S^{un} = \{(\mathbf{p}_t; \mathbf{q}_t), t = 1, \dots, T\}$$

represent the corresponding set of observations that is used in the analysis of unitary rationality.

Given the set S^{un} , unitary rationality means that observed behavior can be rationalized in terms of a single decision maker maximizing a single ('unitary') utility function U . Throughout, we will assume utility functions are continuous, concave, monotonically increasing and non-satiated. Formally, we get the following condition for unitary rationality:

Definition 1 (unitary rationalization, UR) *Let $S^{un} = \{(\mathbf{p}_t; \mathbf{q}_t); t = 1, \dots, T\}$ be a set of observations. A utility function U provides a unitary rationalization (UR) of S^{un} if and only if for each observation $t = 1, \dots, T$: $U(\mathbf{q}_t)$ equals*

$$\max_{\mathbf{z} \in \mathbb{R}_+^n} U(\mathbf{z}) \text{ s.t. } \mathbf{p}'_t \mathbf{z} \leq \mathbf{p}'_t \mathbf{q}_t, \quad (1)$$

with \mathbf{z} representing affordable consumption quantities for the prices \mathbf{p}_t and budget $\mathbf{p}'_t \mathbf{q}_t$.

Varian (1982), based on Afriat (1967), has demonstrated that such a data rationalizing utility function exists if and only if a solution exists for the so-called Afriat inequalities. This is contained in the next result:

Proposition 1 *Let $S^{un} = \{(\mathbf{p}_t; \mathbf{q}_t); t = 1, \dots, T\}$ be a set of observations. The following statements are equivalent:*

(i) *There exists a utility function U that provides a UR of S^{un} ;*

(ii) *For all $s, t \in \{1, \dots, T\}$, there exist numbers $U_t, \lambda_t \in \mathbb{R}_{++}$ that meet the Afriat inequalities*

$$U_s - U_t \leq \lambda_t \mathbf{p}'_t (\mathbf{q}_s - \mathbf{q}_t). \quad (2)$$

In this result, the equivalence between statements (i) and (ii) means that there exists a rationalizing utility function U if and only if the set S^{un} satisfies a number of inequalities defined in the unknowns U_t and λ_t . These last inequalities are commonly referred to as ‘Afriat inequalities’ corresponding to the set S^{un} . Intuitively, these Afriat inequalities allow us to obtain an explicit construction of the utility levels and the marginal utility of income associated with each observation t : they define a utility level U_t and a marginal utility of income λ_t (associated with the observed income $\mathbf{p}'_t \mathbf{q}_t$) for each observed \mathbf{q}_t . We remark that the Afriat inequalities in (2) are linear. Thus, we can use standard linear programming techniques to verify if there exists a unitary rationalization for S .

In what follows, we will consider an extended version of the rationality condition in Proposition 1. The extended condition accounts for optimization errors in the following way: it requires ‘nearly’ optimizing behavior rather than ‘exactly’ optimizing behavior (see Afriat (1973) and Varian (1990) on the usefulness of considering such nearly optimizing behavior in empirical revealed preference analysis). Formally, the extended version of the rationality condition uses $e \in [0, 1]$, and replaces (2) by

$$U_s - U_t \leq \lambda_t \mathbf{p}'_t (\mathbf{q}_s - e \mathbf{q}_t). \quad (3)$$

Clearly, $e = 1$ makes (3) coincide with (2). In our following empirical exercise, we allow for small optimization errors by additionally considering $e = 0.975$ and $e = 0.95$. In general, lower values for e imply weaker unitary rationality conditions. We remark that using (3) instead of (2) preserves the linear nature of the restrictions.

2.2 Collective rationality without bargaining weight restrictions

Our empirical investigation of the collective consumption model will focus on two-member groups or ‘dyads’. Like above, we assume an observed set of T dyad choices, consisting of quantities \mathbf{q}_t and prices \mathbf{p}_t . Now we also use information on the observed within-dyad allocation of the quantities \mathbf{q}_t ; our experimental set-up

allows us to obtain this information (see Section 3). Specifically, for each observed bundle \mathbf{q}_t , we know the individually consumed quantities $\mathbf{q}_t^1, \mathbf{q}_t^2 \in \mathbb{R}_+^N$ such that

$$\mathbf{q}_t = \mathbf{q}_t^1 + \mathbf{q}_t^2, \quad (4)$$

Thus, our analysis of collective rationality will use the set of observations

$$S^{co} = \{(\mathbf{p}_t; \mathbf{q}_t, \mathbf{q}_t^1, \mathbf{q}_t^2), t = 1, \dots, T\}$$

The collective model assumes that the preferences of the two group members, which are defined in terms of the privately consumed quantities, can be represented by utility functions U^1 and U^2 . The collective consumption model (only) assumes Pareto efficient within-group allocations. Formally, we obtain the following condition for collective rationality.

Definition 2 (collective rationalization, CR) *Let $S^{co} = \{(\mathbf{p}_t; \mathbf{q}_t, \mathbf{q}_t^1, \mathbf{q}_t^2), t = 1, \dots, T\}$ be a set of observations. A pair of utility functions U^1 and U^2 provides a collective rationalization (CR) of S^{co} if and only if for each observation $t = 1, \dots, T$: there exists a Pareto weight $\mu_t \in \mathbb{R}_{++}$ such that $U^1(\mathbf{q}_t^1) + \mu_t U^2(\mathbf{q}_t^2)$ equals*

$$\max_{(\mathbf{z}^1, \mathbf{z}^2) \in (\mathbb{R}_+^n)^2} U^1(\mathbf{z}^1) + \mu_t U^2(\mathbf{z}^2) \quad \text{s.t.} \quad \mathbf{p}'_t(\mathbf{z}^1 + \mathbf{z}^2) \leq \mathbf{p}'_t \mathbf{q}_t, \quad (5)$$

with \mathbf{z}^1 and \mathbf{z}^2 representing affordable consumption quantities for the prices \mathbf{p}_t and budget $\mathbf{p}'_t \mathbf{q}_t$.

Thus, the collective consumption model generalizes the unitary model by describing group behavior as maximizing a weighted sum of the individual member utilities. The Pareto weight μ_t then represents the relative ‘bargaining power’ of member 2 (*vis-à-vis* member 1) in observation/situation t . Importantly, this bargaining power may vary depending on the specific observation at hand.

The following result provides a revealed preference characterization of collectively rational behavior.

Proposition 2 *Let $S^{co} = \{(\mathbf{p}_t; \mathbf{q}_t, \mathbf{q}_t^1, \mathbf{q}_t^2), t = 1, \dots, T\}$ be a set of observations. The following statements are equivalent:*

- (i) *There exists a pair of utility functions U^1 and U^2 that provide a CR of S^{co} ;*
- (ii) *For all $s, t \in \{1, \dots, T\}$, $m = 1, 2$ there exist numbers $U_t^m, \lambda_t^m \in \mathbb{R}_{++}$ that meet the collective Afriat inequalities*

$$U_s^m - U_t^m \leq \lambda_t^m \mathbf{p}'_t (\mathbf{q}_s^m - \mathbf{q}_t^m). \quad (6)$$

The interpretation of statement (ii) in this result is similar to the one of statement (ii) in Proposition 1. Just like for individual rationality, collective rationality requires finding a solution for Afriat inequalities. In this case, we get a set of inequalities for each individual member defined in terms of the given personalized quantities and prices. Like before, they allow for an explicit construction of (in casu member-specific) utilities (U_t^m) and marginal utilities of income (λ_t^m). In the collective case, the marginal utilities of income pertain to the income shares of the individual group members associated with the observed intragroup allocation (i.e. $\mathbf{p}'_t \mathbf{q}_t^m$ for each member m).

Similar to before, our following experimental analysis will focus on an extended version of the inequalities (6) that accounts for optimization error, i.e.

$$U_s^m - U_t^m \leq \lambda_t \mathbf{p}'_t (\mathbf{q}_s^m - e \mathbf{q}_t^m) \quad (7)$$

with $e = 1, 0.975$ and 0.95 . The resulting inequalities can again be verified through linear programming techniques.

Two final remarks are in order. Firstly, the above collective consumption model is the so-called egoistic model, in which each member m 's utility function only depends on the own private consumption \mathbf{q}^m . This means that we do not allow (i) for consumption externalities (i.e. U^m does not depend on \mathbf{q}^l , $l \neq m$) or (ii) public consumption within the group. Browning and Chiappori (1998) and Chiappori and Ekeland (2006, 2009) introduced a general collective consumption model that does account for publicly consumed quantities and privately consumed quantities associated with externalities; Cherchye, De Rock and Vermeulen (2007, 2011) developed the revealed preference condition for this general model. We see more refined experiments that specifically focus on consumption externalities and publicly consumed quantities as an interesting avenue for future research. However, we also believe that a simple experiment that focuses on the egoistic model constitutes a useful first step towards considering such more refined settings.

At this point, it is worth indicating that this 'egoistic' specification of the collective model actually encompasses a wider class of member-specific utilities, which model altruism in a specific way: it also includes so-called caring preferences of the individual group members, which depend not only on the member's own (egoistic) utility but also on the other member's utility. Chiappori (1992) argues that the empirical implications of caring preferences are indistinguishable from those of egoistic preferences. As such, while we will not indicate this explicitly in the following discussion, our empirical conclusions for the egoistic model directly

carry over to the (more general) caring model.

A second remark pertains to our maintained assumption that individual utility functions are concave (representing convex preferences). In this respect, Cherchye, de Rock and Vermeulen (2010a) have recently shown that this concavity assumption is actually testable for the egoistic model that we consider here. They established a testable revealed preference condition for the model that applies under general (possibly non-convex) preferences. If we apply this condition to the experimental data presented in Section 4, then we conclude that this last condition cannot be rejected for our dyads under study.⁷ In other words, the egoistic model cannot be rejected if we allow for non-concave utility functions. Or conversely, any rejection of the egoistic model that is reported further on may also be attributed to non-convex individual preferences rather than a rejection of the egoistic model per se.

In this study, we maintain the assumption of concave utilities because it allows us to represent Pareto efficient group behavior as maximizing a weighted sum of individual utilities, with the Pareto weights μ_t indicating the relative bargaining power of the individual group members (see Definition 2). This representation no longer holds if we drop the concavity assumption. As a matter of fact, this weighted sum representation forms the very basis for our methodological contribution that we present next, which develops a revealed preference characterization of collectively rational behavior under restricted Pareto weights μ_t .

2.3 Collective rationality under bargaining weight restrictions

The collective rationality condition discussed above does not impose any restriction on the bargaining weights in different choice situations/observations. If this condition cannot be rejected for a given set of observations, a natural next question asks whether the data also pass a more restricted condition that does impose restrictions on the bargaining weights variation. In this section, we propose a methodology that allows for testing such a condition.

Formally, we consider the following condition of collective rationality under bargaining weight restrictions.

Definition 3 (restricted collective rationalization, a-CR) *Let $S^{co} = \{(\mathbf{p}_t; \mathbf{q}_t, \mathbf{q}_t^1, \mathbf{q}_t^2), t = 1, \dots, T\}$ be a set of observations and consider $a \in \mathbb{R}_+$. A pair of*

⁷For compactness, we do not include these results here. But they are available from the authors upon request.

utility functions U^1 and U^2 provides an a -restricted collective rationalization (a -CR) of S^{co} if and only if for each observation $t = 1, \dots, T$: there exists a constant $C \in \mathbb{R}_+$ and weights $\mu_t \in \mathbb{R}_{++}$ such that $U^1(\mathbf{q}_t^1) + \mu_t U^2(\mathbf{q}_t^2)$ equals

$$\max_{(\mathbf{z}^1, \mathbf{z}^2) \in (\mathbb{R}_+^n)^2} U^1(\mathbf{z}^1) + \mu_t U^2(\mathbf{z}^2) \quad \text{s.t.} \quad \mathbf{p}'_t(\mathbf{z}^1 + \mathbf{z}^2) \leq \mathbf{p}'_t \mathbf{q}_t$$

$$\text{and } \frac{C}{(1+a)} \leq \mu_t \leq C(1+a), \quad (8)$$

with \mathbf{z}^1 and \mathbf{z}^2 representing affordable consumption quantities for the prices \mathbf{p}_t and budget $\mathbf{p}'_t \mathbf{q}_t$.

The only difference with Definition 2 is the additional restriction (8) on the bargaining weights μ_t . In the limiting case when $a = 0$, Definition 3 constraints the bargaining weight to be a constant number C for all observations. For other choices of a , the definition allows for some variation of μ_t , but only in a prespecified (small) range. Finally, taking a arbitrarily large makes Definitions 2 and 3 coincide. As such, Definition 3 allows us to consider a whole spectrum of collective consumption models.

The following proposition presents a revealed preference characterization of collectively rationality under bargaining weight restrictions.⁸

Proposition 3 *Let $S^{co} = \{(\mathbf{p}_t; \mathbf{q}_t, \mathbf{q}_t^1, \mathbf{q}_t^2), t = 1, \dots, T\}$ be a set of observations and consider $a \in \mathbb{R}_+$. The following statements are equivalent:*

- (i) *There exists a pair of utility functions U^1 and U^2 that provide an a -CR of S^{co} ;*
- (ii) *For all $s, t \in \{1, \dots, T\}$, $m = 1, 2$ there exist numbers $U_t^m, \lambda_t^m \in \mathbb{R}_{++}$ that meet the collective Afriat inequalities*

$$U_s^m - U_t^m \leq \lambda_t^m \mathbf{p}'_t (\mathbf{q}_s^m - \mathbf{q}_t^m) \quad (9)$$

$$\text{and } \frac{1}{(1+a)} \leq \frac{\lambda_t^1}{\lambda_t^2} \leq (1+a). \quad (10)$$

The interpretation of the condition in Proposition 3 is directly analogous to the one of Proposition 2. The only difference is the additional restriction (10), which corresponds to the weight restriction (8) in Definition 3. We note that the

⁸Like before, we will replace (9) by (7) to account for optimization error in our empirical exercises.

restriction (10) can be rewritten in linear form as

$$\begin{aligned}\lambda_t^2 &\leq (1+a)\lambda_t^1 \text{ and} \\ \lambda_t^1 &\leq (1+a)\lambda_t^2.\end{aligned}$$

Thus, we can again use linear programming techniques to verify the a -CR condition in Proposition 3.

In our following analysis, the unitary model will serve as a benchmark when evaluating the goodness-of-fit of various collective models. In this respect, it is important to point out that the a -CR condition in Proposition 3 (for any value of a) is not nested with unitary rationality condition in Proposition 1: consistency with the first condition does not imply consistency with the second condition, or vice versa. In our opinion, this nonnestedness conclusion makes it all the more interesting to empirically compare the empirical performance of the unitary model and the collective model (for alternative values of a). We will carry out such a comparison in our empirical analysis.

Example 1 illustrates this nonnestedness result by considering the a -CR condition for $a = 0$.⁹ We believe this case is particularly interesting because it may appear to some that the collective rationalization condition for $a = 0$ is empirically equivalent to the unitary rationalization condition. More precisely, that the constant bargaining weight implies that we can aggregate the individual preferences into a single utility function, and, thus, that we obtain a unitary rationalization. Our Example 1 shows that this reasoning is incorrect. Specifically, it ignores the observed intragroup allocation that is given by the personalized quantities; this intragroup allocation implies additional collective restrictions which are not captured by the unitary rationalization condition.

We emphasize that observing the intragroup allocation (i.e. the set S^{co} contains the personalized quantities \mathbf{q}_t^m) crucially drives this non-nestedness result. Specifically, the result breaks down if the intragroup allocation is not observed (i.e. the set of observation is S^{un} rather than S^{co}). In this case, the collective rationality condition would require that there exists at least one specification of the personalized quantities \mathbf{q}_t^1 and \mathbf{q}_t^2 (with $\mathbf{q}_t^1 + \mathbf{q}_t^2 = \mathbf{q}_t$) such that the correspondingly defined set $\{(\mathbf{p}_t; \mathbf{q}_t, \mathbf{q}_t^1, \mathbf{q}_t^2), t = 1, \dots, T\}$ satisfies statement (ii) in Proposition 3.¹⁰ One can verify that, for $a = 0$, this collective rationality condition is exactly equivalent

⁹Similar (but mathematically less elegant) examples can be constructed to illustrate nonnestedness for $a \neq 0$.

¹⁰Cherchye, De Rock and Vermeulen (2007, 2011) provide an extensive discussion (including formal statements) of collective rationality conditions if the intragroup allocation is not observed.

to the unitary rationality condition.

Example 1 *Suppose a first situation with 2 observations and 2 goods. The set S^{un} includes the following aggregate quantities and prices*

$$\begin{aligned}\mathbf{q}_1 &= (2, 1)', \mathbf{p}_1 = (2, 1)', \\ \mathbf{q}_2 &= (1, 2)', \mathbf{p}_2 = (1, 2)'.\end{aligned}$$

The set S^{co} additionally contains the personalized quantities

$$\begin{aligned}\mathbf{q}_1^1 &= (2, 0)' \text{ and } \mathbf{q}_1^2 = (0, 1)'; \\ \mathbf{q}_2^1 &= (1, 0)' \text{ and } \mathbf{q}_2^2 = (0, 2)'.\end{aligned}$$

Using linear programming it is now easy to verify that the set S^{co} meets the 0-CR condition in Proposition 3 while the set S^{un} does not meet the UR condition in Proposition 1.

Next, suppose a second situation with again 2 goods and 2 observations. In this case, the set S^{un} includes

$$\begin{aligned}\mathbf{q}_1 &= (2, 1)', \mathbf{p}_1 = (1, 2)', \\ \mathbf{q}_2 &= (1, 2)', \mathbf{p}_2 = (2, 1)';\end{aligned}$$

while the set S^{co} additionally contains

$$\begin{aligned}\mathbf{q}_1^1 &= (0, 1)' \text{ and } \mathbf{q}_1^2 = (2, 0)'; \\ \mathbf{q}_2^1 &= (1, 0)' \text{ and } \mathbf{q}_2^2 = (0, 2)'.\end{aligned}$$

Then one can verify that the set S^{un} satisfies the UR condition, but the set S^{co} does not meet the 0-CR condition.

As a side result, Example 1 also shows that the within-group income distribution may remain constant even if the distribution of the bargaining power varies. Specifically, as we indicated above, the first data set is consistent with a constant bargaining power of the individual group members in the two choice observations. However, the relative income share of the two members varies over the two observations. Similarly, we could construct a data set characterized by a constant income distribution but a varying bargaining power. This shows that, in general, a constant income distribution does not imply a constant bargaining power, and vice versa.

3 Experimental design

Participants in our experiment were 102 undergraduate students (53 women). Ages ranged from 18 years to 25 years (mean value = 21.02; standard deviation = 1.72). As we wanted to analyze collective choice behavior, both men and women were asked to sign up for an experimental session together with either a male or a female friend or a romantic partner. This led to a sample containing four types of dyads, namely, male dyads or two male friends (12 in total), female dyads or two female friends (14 in total), mixed dyads or one male and one female friend that do not bound romantically (13 in total), and romantic dyads or one man and one woman who were in a romantic relationship together (12 in total).¹¹ Given the small sample sizes for the different dyad types, we will only report test results for the full sample in what follows. Test results for specific dyad types are available from the authors upon request.

Participants were scheduled to come to the laboratory in groups of eight (i.e., four dyads). Each participant was assigned a seat in a partially enclosed cubicle, and worked individually for the main part of the session. Dyads were asked to engage in one experimental task together. Participants were rewarded with money and with a commodity bundle for their cooperation. Each dyad received money and a commodity bundle with a combined value of € 20.

Our experiment is similar in design to the one of Harbaugh, Krause and Berry (2001), who used a revealed preference methodology to investigate individual rationality for young children. Upon entering the laboratory, participants were given the opportunity to taste small quantities of red wine, orange juice, and M&Ms (i.e. a type of chocolate candy). They were truthfully told that they would be making consumption decisions with respect to these three commodities later on, and that we wanted them to familiarize themselves with the commodities. Participants were then presented with 9 choice problems, i.e. $T = 9$. Each choice consisted of the three commodities red wine, orange juice, and M&Ms, i.e. $N = 3$. We selected these goods following previous studies that used revealed preference methodology for analyzing individual rationality (see, for example, Sippel (1997) and Harbaugh, Krause, and Berry (2001)).

Each choice problem was characterized by a different price regime; the prices of the three commodities are shown in Table 1. The participants were asked to indicate how much they wanted to obtain from each commodity, given that the total budget they could allocate to the three commodities was € 10. In order to

¹¹Homosexual nor lesbian dyads were included in the study.

avoid indivisibility issues, we created very small units (i.e. a unit of red wine is 1 centiliter, a unit of orange juice is 3 centiliters, and a unit of M&Ms is 5 grams); in addition, participants had the option to select non-integer quantities. We note that the price-income regimes in our experiment imply a high power of our rationality tests (i.e. a high probability of detecting irrational behavior), essentially because there is no variation of income (€ 10) but a lot of variation of prices. For example, Blundell, Browning, and Crawford (2003) apply a similar idea in their ‘maximum power sequential path’ procedure for maximizing the power of their nonparametric tests. We will return to the power of our tests in the next section.

Participants were asked to make the 9 allocation decisions twice: once individually and once together with their friend. The order in which both sets of decisions were to be made was counterbalanced: one half of the dyads first made the decisions individually and only afterwards collectively, whereas the other half of the dyads first made the decisions collectively and only afterwards individually; changing the order in this way did not yield significantly different results in terms of (individual or collective) rationality. Table 2 presents summary information on the budget shares corresponding with the individuals’ and the dyads’ choices under the 9 price regimes; this expenditure information also allows for reconstructing the corresponding (mean) quantities that have been chosen under the different price regimes in Table 1.

In case of collective decision-making, participants were asked to indicate for each of the three commodities which percentage of their demand was intended for each individual. This provides the personalized quantity information that we use for the collective rationality tests discussed above.

The decision problems participants were faced with were supposed to mimic real life difficulties that both individual consumers and groups often encounter when having to pick their optimal commodity bundles out of the available budget sets. To enhance the external validity of our study, participants were told that, when all experimental sessions were over (maximum two weeks after they participated), they would actually receive one of the commodity bundles they had put together. They were also told that they would be informed through E-mail about where and when to pick it up. For practical reasons, we picked this bundle randomly from the set of decisions that participants had made collectively (and we thus ignored the individually chosen bundles), although they were not informed of this beforehand. The knowledge that each choice ostensibly had the same chance of actually being implemented was supposed to give economic significance to otherwise merely hypothetical decisions, thus providing participants with an incentive

for making choices that truly represented their preferences.

As making the allocation decisions required a considerable amount of calculation (multiplying prices and demand for each commodity and adding up to check whether the budget is exhausted), participants were encouraged to use a calculator to check their decisions. Participants could also spend as much time as they liked on their decisions and were free to compare, reconsider, and correct previous choices. As such we maximally avoid dynamic and intertemporal issues (such as learning, punishment strategies, etc.), which falls in line with the static nature of the collective model under consideration. When the participants felt that the decisions they had made represented their actual preferences, the experimenter provided them with the instructions for the next task. In Appendix A, we include the instructions that were handed out to the participants.

Before considering our actual test results, it is interesting to have a closer look at how dyad members shared the income that was given to them. Table 3 provides results for the 9 choice problems in our experiment. To define the income shares, we make use of the personalized quantity information that has been reported; using the notation of Section 2, each member m 's income share is defined as $\mathbf{p}'_t \mathbf{q}_t^m$. Table 3 gives summary information on income share differences in relative terms, i.e. absolute differences between individual income shares divided by the total dyad income (€ 10). It reports mean values, as well as corresponding standard deviations and maximum values. In addition, for each choice problem, the column 'equal split' provides the percentage of dyads that apply equal sharing.

We find that dyads opt for equal sharing in about 30% to 40% of all choices. This means that there is an unequal distribution of resources in about two thirds of all choice situations. We believe this is quite substantial, especially when taking into account the laboratory nature of our experiment. In fact, in some cases the income inequality is very large: in one instance, the (maximum) difference between members' income shares amount to no less than 80% of total income. Furthermore, the mean values of relative income differences are rather pronounced and there is considerable variation across choice observations (see the standard deviations). All this makes us believe that this is a useful data set for demonstrating the practical usefulness of the revealed preference methodology introduced above.

As a concluding remark, we indicate that the original experiment also included an additional group of dyads that were confronted with almost identical decision problems (i.e. they had to state their demand for red wine, orange juice, and M&Ms, given the same relative price variations as presented in Table 1 and a budget of €10), but with the additional option of receiving in cash any amount of

the budget they wanted to in each decision situation; the price of this additional ‘cash’ commodity equals 1 for all choice problems. The test results for this ‘4-commodities’ group of dyads turned out to be qualitatively similar to the results for the group that did not have the cash option. Therefore, to keep our discussion compact, we will not incorporate the test results for the 4-commodities group in the following section. These results are available upon simple request.

In fact, the cash option for the 4-commodities group also served as a validity check of our experimental design. Specifically, if the dyads found the experiment not appealing (because of the commodities and/or their prices), then the cashed amount would equal their total budget in each observation; this would guarantee that they received the money with certainty. Because participants effectively spent a large share of their total budget on the three original commodities (rather than keeping it as cash), we conclude that our design does pass this validity test.¹²

Table 1 about here

Table 2 about here

Table 3 about here

4 Test results

We will first report on the rationality of the individual choices. Indeed, individually rational behavior is a prerequisite for collectively rational behavior. As a result, a first test for rational collective behavior is whether the unitary model (which assumes a single decision maker) adequately describes the observed individual choice behavior. Subsequently, we will consider the empirical goodness-of-fit of the unitary model for describing the collective choices. This will set the benchmark for the collective consumption model, which will be considered in the last section.

We will consider the goodness-of-fit as well as the discriminatory power of all models under evaluation. We believe both aspects are important when comparing the empirical performance of alternative consumption models. In particular, we address the adequacy of different models by following the proposal of Beatty and Crawford (2011). These authors propose a ‘predictive success’ measure to evaluate the overall empirical performance of specific behavioral models in the context of revealed preference analysis.

¹²On average, participants spent 65% of their total budget on the three commodities, with a standard deviation of 27% .

More specifically, Beatty and Crawford’s measure of predictive success simultaneously accounts for the two empirical performance aspects mentioned above. Goodness-of-fit (‘fit’) is measured as the percentage of study subjects (individuals or groups) passing the rationality test for the model under study. Next, discriminatory power (‘power’) is measured as the probability that the test detects behavior that is irrational; to compute this power, we will model irrational behavior as random behavior (see below). Fit and power values lie between 0 and 1. Predictive success is then defined as

$$\text{predictive success} = \text{power} - (1 - \text{fit}).$$

The value of this measure always lies between -1 and 1 . A value close to 1 indicates a model with approximately 100% power and 100% fit, i.e. the best possible scenario. This means that (almost) all data pass the rationality test, even though the test effectively detects (almost) any deviating (i.e. random or irrational) behavior. By contrast, a value close to -1 implies a model with approximately 0% power and 0% fit, i.e. the worst possible scenario. In this case, the test effectively allows for (almost) any observed behavior and yet the data fail to pass. Finally, a value of 0 corresponds to a model with a rejection rate for the observed behavior ($= 1 - \text{fit}$) that exactly equals the expected rejection rate if behavior were random ($= \text{power}$). Essentially, this means that the rationality test does not allow for distinguishing observed behavior from random behavior. We believe this 0 value provides a useful benchmark value. A ‘good’ model must at least have a predictive success rate that is positive. In general, a higher predictive success rate reveals a better empirical performance of a given model.

At this point, it is worth indicating that the predictive success measure defined above actually assigns an equal weight to discriminatory power and goodness-of-fit. This equal weighting may seem arbitrary to some. Interestingly, however, Beatty and Crawford (2011) show that this weighting scheme has an interesting axiomatic characterization.¹³ We believe this provides a convincing theoretical foundation for our focus on the (equally weighted) predictive success measure as it was originally presented by Beatty and Crawford.

¹³Specifically, by using an original theorem of Selten (1991), Beatty and Crawford (2011) show that their (equally weighted) predictive success measure can be characterized by three axioms: monotonicity, equivalence and aggregability. For brevity, we do not give a formal definition of these axioms here, but refer to the study of Beatty and Crawford for a detailed discussion. These authors also provide an intuitive explanation/motivation for these axioms in a revealed preference setting such as ours.

4.1 Individual rationality test

To set the stage, we discuss the test results for individual rationality, based on the rationality condition in Proposition 1. As explained before, we account for optimization error by focusing on the extended version of the Afriat inequalities in (3). In these inequalities, lower values of e account for greater optimization error. In our application, we consider the models with $e = 1$ ('100% optimization'), 0.975 ('97.5% optimization') and 0.95 ('95% optimization').

Table 4 presents the results. Let us first consider the goodness-of-fit values for the three models that we consider. We find that all models provide an adequate description of observed behavior. Not surprisingly, lower values for e imply higher pass rates. But even the 100% optimization model does rationalize the behavior of slightly more than 90% of the individuals under consideration. The fit of the other models amounts to approximately 100%. These results are consistent with those of other, similar experiments on individual rationality (e.g. Harbaugh, Krause and Berry, 2001, and Andreoni and Miller, 2002).

Let us then consider power. As indicated above, we measure power as the probability of detecting random behavior. We model random behavior using the bootstrap method for panel data as described by Andreoni and Miller (2002) and applied by Harbaugh, Krause and Berry (2001) within a similar experimental context.¹⁴ The method essentially mimics random behavior for each price regime (or budget) by drawing randomly from the observed set of choices under that price regime (i.e. 102 choices under 9 different price regimes). This gives information on the expected distribution of violations under random choice, while incorporating information on the participants' actual choices. All bootstrap results reported in this paper are based on Monte Carlo-type simulations that include 50000 iterations.

Table 4 shows that random behavior leads to rejections of the 100% optimization model in 82% of the cases. This confirms that our experimental design leads to a powerful test of the model. Power decreases rather drastically when e decreases. For example, the 95% optimization model rejects rationality of random behavior in only about 60% of the cases.

¹⁴This bootstrap method is similar to the randomization method proposed by Bronars (1987), which has also been used frequently in the literature. The mere difference is that 'random' choices (for each price regime) are drawn from the observed distribution whereas Bronars randomly draws from the uniform distribution (which may significantly differ from the observed distribution). We refer to Andreoni and Harbaugh (2006) for a detailed discussion of the strengths and weaknesses of alternative randomization procedures, and corresponding power measures, that have been used within a nonparametric context.

Finally, Table 4 gives the predictive success rates for the three individual rationality models that we consider. All models have a predictive success rate that is substantially above 0, which is comforting. Next, we find that the 100% optimization model substantially outperforms the 97.5% and 95% optimization models. The explanation is that discriminatory power is substantially higher for the former model than for the latter two models, even though the goodness-of-fit results show an (albeit much less pronounced) opposite pattern.

Our overall conclusion is that the individual rationality model cannot be rejected for our specific choice setting. The ‘best’ model is the 100% optimization model, which has a predictive success rate of about 75%. Next, on the basis of the results in Table 4, we can reasonably assume that participants in our experiment satisfy individual rationality. This provides a useful motivation for our following exercises, which investigate rationality of the same individuals in collective choice settings.

Table 4 about here

4.2 Unitary rationality test for dyads

We first study rationality of collective choices in terms of the unitary model. Like before, we consider the condition in Proposition 1, and account for optimization error by focusing on the Afriat inequalities in (3); we again use $e = 1$ (‘100% optimization’), 0.975 (‘97.5% optimization’) and 0.95 (‘95% optimization’). We compute the power of the different models by using a similar bootstrap method as before. The only difference is that, for each price regime, we now mimic random behavior by drawing randomly from the observed 51 collective choices.

Table 5 reports our findings. Overall, these results are closely similar to the ones in Table 4. First, goodness-of-fit is rather high for the three model specifications. Next, power is substantial for the 100% optimization model, but decreases considerably for the 97.5% and 95% optimization models. All this yields that the 100% optimization model has the best predictive success rate and, thus, can be regarded as the ‘best’ unitary model for describing the observed collective behavior.

We conclude that the unitary model does a good job in describing observed dyad behavior in our experiment. For example, the predictive success rate of the 100% optimization model is about 72%. Next, we compare the empirical performance of this model with the one of the collective model with varying bargaining weight restrictions.

Table 5 about here

4.3 Collective rationality test for dyads

Let us then consider the empirical performance of the collective model. As an introductory note, we recall that our results for the individual choices provide do not reject the hypothesis of individual rational behavior (see our discussion of Table 4). This suggests that we can safely maintain individual rationality as an assumption in our following tests of collective rationality. Following this interpretation, the collective rationality test results can thus be seen as checking the validity of the Pareto efficiency hypothesis for the group (in casu dyad) decision process. As discussed above, Pareto efficiency is effectively the distinguishing hypothesis of the collective consumption model.

We test the collective rationality condition in Proposition 3 for the observed dyad choices; and we study varying restrictions for the bargaining weights by considering multiple values for a situated between 0 and arbitrarily large (or *free*; see Table 6). We recall that $a = 0$ corresponds to the collective model with constant bargaining weights; while $a = \textit{free}$ implies no weight restriction at all (i.e. the conditions in Propositions 2 and 3 coincide). Generally, lower values for a correspond to stronger weight restrictions (see Definition 3). Next, we again account for optimization error by using the Afriat inequalities in (7) with $e = 1$ ('100% optimization'), 0.975 ('97.5% optimization') and 0.95 ('95% optimization'). Finally, power is measured in the same way as for the unitary model.

Table 6 presents our results. We first consider goodness-of-fit. The model's goodness-of-fit increases substantially when allowing for some optimization error. For all values of a , the 97.5% optimization model significantly outperforms the 100% optimization model: the former model has a fit above 90% (and often close to 100%) for any specification of a , while the fit of the latter model varies between 62% and 85% (depending on a). Next, we find that the 95% optimization model (only) slightly improves upon the 97.5% optimization model, which is obvious given that the latter has already a fit above 90%. Lastly, we observe that goodness-of-fit improves for higher a . This is not unexpected as higher values for a imply less stringent empirical restrictions.

As for power, we find that all models perform reasonably well. As expected, power decreases with the value of e . However, this power decrease as a function of e is generally less pronounced than for the unitary model (compare with Table 5). We also observe that power decreases when a increases, which is again not sur-

prising. Interestingly, for all values of e the discriminatory power of the collective model (for any value of a) substantially exceeds the power for the unitary model. This is due to the experimental design that allows us to observe the personalized quantities. In our opinion, this suggests the empirical usefulness of our methodology to assess several specifications of the collective consumption model by means of experimental data. Such an assessment effectively allows us to meaningfully analyze the empirical validity of the different models.

Next, a most notable observation is that many specifications of the collective model (often substantially) outperform the ‘best’ unitary model in terms of predictive success. A lot of specifications have a predictive success of no less than 85%. Generally, our results suggest that the ‘best’ specifications of the collective model allow for a limited variation of the bargaining weights (i.e. low value for a) in combination with a small optimization error (i.e. $e = 0.975$). These specifications systematically combine a high fit value (about 95%) with a high power value (about 90%). In our view, these favorable results for specifications with restricted (but not constant) bargaining weights are particularly appealing. Indeed, as we discussed in the introductory section, such specifications may effectively provide a more realistic description of group consumption behavior than the model specification that does not include such restrictions.

As a final exercise, we have investigated the possible impact of some specific dyad features on our collective rationality results. Specifically, we conducted an analysis of variance (ANOVA) that included the following information for the participants in our experiment: duration of the period (months) that friends/partners know each other, different indicators of the degree of friendship, and alternative variables capturing the way in which friends/partners interact with each other. However, we did not detect any significant effects. A possible explanation may be the rather small size of our sample.

Table 6 about here

4.4 Towards household consumption data

Next, we examine the relevance of our findings for the analysis of household consumption behavior. Our study provides a useful complement of existing studies that assess the empirical validity of the collective model by using a parametric specification and/or ‘real life’ data. Deviating from these studies, we address the same question by using nonparametric (revealed preference) analytical tools in combination with experimental data. As such, our tests are ‘pure’ in that they

avoid (i) a (non-verifiable) parametric structure that is imposed on the model, and (ii) preference homogeneity assumptions and data measurement problems that are specific to ‘real life’ data.

We recall from our discussion in Section 2 that our analysis has focused on the egoistic model, in which each member’s utility function only depends on the own private consumption (excluding consumption externalities and public consumption). The reason is that empirical applications of the collective model based on real life data mostly assume this model. In this respect, we note that our first experimental test involved an unsophisticated consumption setting, with a very limited number of commodities and a low budget. Interestingly, our findings suggest that the egoistic model effectively constitutes a useful tool for describing collective choice behavior in such a simple setting: we obtain reasonably high predictive success rates for collective rationality models that allow for a limited variation of the bargaining weights. This can be seen as a minimal validity check for using the egoistic model in more sophisticated (real life) settings. Of course, the nature of our experiment only allows us to draw suggestive conclusions in this respect.

One final remark applies to using the newly proposed bargaining weight restrictions in the context of household data. It follows from our discussion in Section 2 that the practical application of such restrictions (through linear programming) actually requires that the personalized consumption quantities (\mathbf{q}_t^m for each member m) are observed.¹⁵ This may be problematic in a household context: we argued in the Introduction that household data sets usually only contain information on the aggregate household consumption and not on the individual consumption. In this respect, however, we must add that data sets with personalized quantity information are increasingly available in the literature (see, for example, Bonke and Browning (2006), Browning and Gørtz (2006) and Cherchye, De Rock and Vermeulen (2010b)). For such data sets our methodology is directly applicable, which may thus obtain a vigorous revealed preference analysis of household consumption behavior in terms of the collective model.

¹⁵More precisely, we can verify that the Afriat inequalities in Propositions 2 and 3 become nonlinear in unobservables if such personalized quantity information is not observed (i.e. the set of observation is S^{un} rather than S^{co}). And it is well-known that checking such nonlinear conditions is generally difficult.

5 Summary and conclusion

We have provided a first revealed preference test of the collective consumption model on the basis of experimental data. By using revealed preference methodology and experimental data, we avoid the usual problems associated with parametric tests (e.g. non-verifiable parametric structure) and the use of ‘real life’ data sets (e.g. preference heterogeneity). We have proposed a testing methodology that allows us to restrict the variation of the bargaining weights across choice observations in the revealed preference analysis.

Our empirical analysis has focused on the goodness-of-fit as well as on the discriminatory power of the unitary model and alternative specifications of the collective model (with varying bargaining weight restrictions). Adopting a recent proposal of Beatty and Crawford (2011), we have used a ‘predictive success’ measure to evaluate the overall empirical performance of a specific behavioral model. This measure also enabled us to compare this empirical performance for alternative model specifications.

Our results indicate that including bargaining weight restrictions may (often substantially) increase the discriminatory power of the revealed preference tests. In our opinion, this is a most interesting observation, as low power is a frequently cited concern for revealed preference analysis. Bargaining weight restrictions imply additional structure for the decision process that can easily be motivated on a priori grounds. Indeed, as indicated in the Introduction, it often seems reasonable to consider very large bargaining power shifts as unrealistic. As such, we believe our methodology to restrict bargaining weight variation provides a natural way to increase the power of the revealed preference analysis of collective consumption models.

At a more specific level, our results suggest that the choices made in our experiment are best described by a collective model that allows for a limited variation of the bargaining weights. This model specification has higher predictive success than the specification with constant bargaining weights and the specification with unlimited variation of the bargaining weights. In our opinion, this is a useful finding as the model with restricted (but not constant) bargaining weights may effectively be considered as a realistic model of group consumption behavior. Interestingly, this collective model also outperforms the unitary model in terms of predictive success. In particular, the model is characterized by substantially more discriminatory power than the unitary model, while the difference in terms of goodness-of-fit is much less pronounced.

At a more general level, our analysis provides further empirical support for considering non-unitary models to describe the behavior of multi-person groups (such as households). In particular, they motivate the use of group consumption models that explicitly recognize the individual preferences within the group. Next, our study demonstrates the usefulness of experimental analysis of group consumption behavior. In particular, such an analysis easily allows for obtaining information on consumption quantities for the individual group members, which enhances the power of the analysis.

As for follow-up research, we believe future work may fruitfully focus on the experimental analysis of group consumption processes in more complicated consumption settings. In this first study, we considered a rather simple choice setting with only two group members, three commodities and a very low joint budget. It would be interesting to contrast our findings here with test results that focus on more complicated settings that focus on more group members, many goods and/or a larger group budget. In doing so, one can also conceive settings that involve publicly consumed goods. For example, one can then investigate whether group decisions on these public goods are consistent with the ‘cooperative’ Pareto efficiency concept or rather the ‘noncooperative’ Nash equilibrium concept.¹⁶ In a similar vein, future experimental studies may focus on consumption externalities associated with privately consumed quantities.

In this respect, one may also analyze alternative (non-consensual) group consumption models that account for independent individual decision making (see, for example, Grossbard-Shechtman (1984, 2003) and Grossbard (2010)). This would allow one to assess (and compare) the empirical performance of different models proposed in the literature on ‘New Home Economics’. We believe that our study illustrates the potential of using revealed preference methodology in combination with experimental data for addressing this type of questions, which can provide additional insight into the appropriate modeling of the consumption behavior of multi-person groups.

¹⁶Lechene and Preston (2005, 2010) and Browning, Chiappori and Lechene (2010) provide recent discussions on modeling noncooperative household consumption behavior, which may imply deviations from Pareto efficiency (in the case of public goods). Cherchye, Demuynck and De Rock (2011) established the revealed preference characterization of the noncooperative model of household consumption.

Appendix A

In this appendix, we add the instructions that were given to the participants in the experiment and that are not explicitly taken up in the main text:

Introduction and product-tasting

‘In a moment, you will be asked to make a series of choices regarding the three products standing on the desk in front of you: wine, orange juice, and M&Ms.

In order to enable you to make these choices in an informed way, you are given the opportunity to taste these three products (wine, orange juice, and M&Ms) right now. You can consume everything if you want to.

You will be asked to make the choices partly on your own and partly together with the person who accompanied you to the lab.

It is in your own and in your joint interest to make these choices as truthfully as possible, that is, to make the same choices as you would in real life. This is because one of your choices will be randomly selected for you to take home.

Because of practical reasons, you will not receive your choice right away. You will however be invited to pick up your choice afterwards. Each choice will consist of a product package (of wine, orange juice, and M&Ms) that is worth €10.’

Decision-making

‘In the following, you are asked to make a series of choices. Each time, relative prices of three different products (wine, orange juice, and M&Ms) are given. It is up to you to decide how much you are willing to spend on each product, given these relative price variations. Each time, your budget amounts to € 10 or 1000 euro cents. Please make each choice as truthfully as possible, that is, as you would in real life, as one of your choices will be randomly selected for you to take home.

If you want to, you can use the computer’s calculator. You are also free to use the back of the questionnaire for scrap paper if you want to.’

Appendix B

We only prove Proposition 3. The proof of Proposition 2 pertains to a specific case of the general case considered in Proposition 3.

(i) \Rightarrow (ii). Under condition (i), we have that each $(\mathbf{q}_t^1, \mathbf{q}_t^2)$ solves the problem

$$\max_{(\mathbf{z}^1, \mathbf{z}^2) \in (\mathbb{R}_+^n)^2} U^1(\mathbf{z}^1) + \mu_t U^2(\mathbf{z}^2) \text{ s.t. } \mathbf{p}'_t(\mathbf{z}^1 + \mathbf{z}^2) \leq \mathbf{p}'_t \mathbf{q}_t.$$

Moreover there exists a $C \in \mathbb{R}_+$ such that $\frac{C}{(1+a)} \leq \mu_t \leq C(1+a)$. Given concavity, the functions U^1, U^2 are subdifferentiable, which carries over to their weighted sum $U^1 + \mu_t U^2$.¹⁷ An optimal solution to the above maximization problem must therefore satisfy (for η_t the Lagrange multiplier associated with the budget constraint)

$$U_{\mathbf{q}_t^1}^1 \leq \eta_t \mathbf{p}_t \text{ and } \mu_t U_{\mathbf{q}_t^2}^2 \leq \eta_t \mathbf{p}_t,$$

for $U_{\mathbf{q}_t^m}^m$ a subgradient of the function U^m defined for the vector \mathbf{q}^m and evaluated at \mathbf{q}_t^m (with $m = 1, 2$). Taking $\lambda_t^1 = \frac{\eta_t}{C}$ and $\lambda_t^2 = \frac{\eta_t}{\mu_t}$ gives us

$$\frac{1}{C} U_{\mathbf{q}_t^1}^1 \leq \lambda_t^1 \mathbf{p}_t \text{ and } U_{\mathbf{q}_t^2}^2 \leq \lambda_t^2 \mathbf{p}_t. \quad (11)$$

Moreover, by construction we have $\frac{\lambda_t^1}{\lambda_t^2} = \frac{\mu_t}{C}$, which shows that (10) holds.

Next, concavity of the functions U^1, U^2 implies for each $m = 1, 2$

$$U^m(\mathbf{q}_s^m) - U^m(\mathbf{q}_t^m) \leq U_{\mathbf{q}_t^m}^m(\mathbf{q}_s^m - \mathbf{q}_t^m). \quad (12)$$

Substituting (11) in (12) and setting $U_k^1 = \frac{1}{C} U^1(\mathbf{q}_k^1)$ and $U_k^2 = U^2(\mathbf{q}_k^2)$ (with $k = s, t$) obtains (9), which shows that condition (ii) holds.

(ii) \Rightarrow (i). Under condition (ii), for any $(\mathbf{z}^1, \mathbf{z}^2)$ such that $\mathbf{p}'_t(\mathbf{z}^1 + \mathbf{z}^2) \leq \mathbf{p}'_t \mathbf{q}_t$ we can define for all $m = 1, 2$

$$U^m(\mathbf{z}^m) = \min_{s \in \{1, \dots, T\}} [U_s^m + \lambda_s^m(\mathbf{p}'_s \mathbf{z}^m - \mathbf{p}'_s \mathbf{q}_s^m)]. \quad (13)$$

Varian (1982) proves that $U^m(\mathbf{q}_t^m) = U_t^m$. Next, given $\mu_t \in \mathbb{R}_{++}$, we have that

$$U^1(\mathbf{z}^1) + \mu_t U^2(\mathbf{z}^2) \leq U_t^1 + \lambda_t^1(\mathbf{p}'_t \mathbf{z}^1 - \mathbf{p}'_t \mathbf{q}_t^1) + \mu_t [U_t^2 + \lambda_t^2(\mathbf{p}'_t \mathbf{z}^2 - \mathbf{p}'_t \mathbf{q}_t^2)].$$

Without losing generality, we concentrate on $\mu_t = \lambda_t^1 / \lambda_t^2$, which obtains

$$U^1(\mathbf{z}^1) + \mu_t U^1(\mathbf{z}^1) \leq U_t^1 + \mu_t U_t^2 + \lambda_t^1 [\mathbf{p}'_t(\mathbf{z}^1 + \mathbf{z}^2) - \mathbf{p}'_t(\mathbf{q}_t^1 + \mathbf{q}_t^2)].$$

¹⁷To be precise, $-U^1, -U^2$ is convex and therefore subdifferentiable. This, of course, does not affect our argument.

Since $\mathbf{p}'_t(\mathbf{z}^1 + \mathbf{z}^2) \leq \mathbf{p}'_t\mathbf{q}_t$, we thus have

$$U^1(\mathbf{z}^1) + \mu_t U^1(\mathbf{z}^2) \leq U_t^1 + \mu_t U_t^2 = U^1(\mathbf{q}_t^1) + \mu_t U^2(\mathbf{q}_t^2),$$

which proves that $(\mathbf{q}_t^1, \mathbf{q}_t^2)$ maximizes $U^1(\mathbf{z}^1) + \mu_t U^2(\mathbf{z}^2)$ subject to $\mathbf{p}'_t(\mathbf{z}^1 + \mathbf{z}^2) \leq \mathbf{p}'_t\mathbf{q}_t$. We conclude that the functions U^1, U^2 in (13) provide a collective rationalization of the set S . Varian (1982) shows that these utility functions are well-behaved. Moreover, by taking $C = 1$ we also obtain that (8) is satisfied since $\mu_t = \lambda_t^1/\lambda_t^2$.

References

- [1] Afriat, S. (1967), The construction of utility functions from expenditure data, *International Economic Review*, 8, 67-77.
- [2] Afriat, S. (1973), On a System of Inequalities in Demand Analysis: an Extension of the Classical Method, *International Economic Review*, 14, 460-472.
- [3] Andreoni, J. and W.T. Harbaugh (2006), Power indices for revealed preference tests, *University of Wisconsin-Madison Department of Economics Working Paper*, 2005-10, version March 2006.
- [4] Andreoni, J. and J. Miller (2002), Giving According to GARP: An experimental test of the consistency of preferences for altruism, *Econometrica* 70, 737-753.
- [5] Apps, P. and R. Rees (1988), Taxation and the household, *Journal of Public Economics* 35, 355-369.
- [6] Apps, P. and R. Rees (2009), Public economics and the household, *Cambridge University Press*.
- [7] Beatty, J. and I. Crawford (2011), How Demanding is the Revealed Preference Approach to Demand?, *American Economic Review*, forthcoming.
- [8] Becker, G. (1973), A theory of marriage: Part I, *Journal of Political Economy* 81, 813-846.
- [9] Blundell, R.W, M. Browning and I.A. Crawford (2003), Nonparametric Engel curves and revealed preferences, *Econometrica* 71, 205-240.
- [10] Bonke, J. and M. Browning (2006), Allocation within the household: direct survey evidence, Working paper 286, Department of Economics, University of Oxford.

- [11] Bronars, S. (1987), The power of nonparametric tests of preference maximization, *Econometrica*, 55, 693-698.
- [12] Browning, M. and P.-A. Chiappori (1998), Efficient intra-household allocations: a general characterization and empirical tests, *Econometrica* 66, 1241-1278.
- [13] Browning, M., P.-A. Chiappori and V. Lechene (2006), Collective and Unitary Models: A Clarification, *Review of Economics of the Household* 4, 5-14.
- [14] Browning, M, P.-A. Chiappori and V. Lechene (2010), Distributional effects in household models: separate spheres and income pooling, *Economic Journal*, 120, 786-799.
- [15] Browning, M., P.-A. Chiappori and A. Lewbel (2006), Estimating consumption economies of scale, adult equivalence scales, and household bargaining power, Boston College Working Paper WP588, Department of Economics, Boston College, version 2006.
- [16] Browning, M. and M. Gørtz (2006), Spending time and money within the household, Economics Series Working Papers 288, University of Oxford, Department of Economics.
- [17] Cherchye, L., T. Demuynck and B. De Rock (2011), Revealed preference analysis of noncooperative household consumption, *Economic Journal*, forthcoming.
- [18] Cherchye, L., B. De Rock and F. Vermeulen (2007), The collective model of household consumption: a nonparametric characterization, *Econometrica*, 75, 553-574.
- [19] Cherchye, L., B. De Rock, and F. Vermeulen (2009), Opening the black box of intra-household decision-making: theory and non-parametric empirical tests of general collective consumption models, *Journal of Political Economy*, 117, 1074-1104.
- [20] Cherchye, L., B. De Rock and F. Vermeulen (2010a), An Afriat theorem for the collective model of household consumption, *Journal of Economic Theory* 145, 1142-1163.
- [21] Cherchye, L., B. De Rock and F. Vermeulen (2010b), Married with children. A collective labor supply model with detailed time use and intrahousehold expenditure information. CentER Discussion paper, Tilburg University.
- [22] Cherchye, L., B. De Rock and F. Vermeulen (2011), The revealed preference approach to collective consumption behavior: testing and sharing rule recovery, *Review of Economic Studies* 78, 176-198.

- [23] Cherchye, L. and F. Vermeulen (2008), Nonparametric analysis of household labor supply: goodness-of-fit and power of the unitary and collective model, *Review of Economics and Statistics* 90, 267-274.
- [24] Chiappori, P.-A. (1988), Rational household labor supply, *Econometrica* 56, 63-89.
- [25] Chiappori, P.-A. (1992), Collective labor supply and welfare, *Journal of Political Economy* 100, 437-467.
- [26] Chiappori, P.-A. and I. Ekeland (2006), The micro economics of group behavior: general characterization, *Journal of Economic Theory*, 130, 1-26.
- [27] Chiappori, P.-A. and I. Ekeland (2009), The micro economics of efficient group behavior: identification, *Econometrica*, 77, 763-799.
- [28] Chiappori, P.-A., B. Fortin and G. Lacroix (2002), Household labor supply, sharing rule, and the marriage marker, *Journal of Political Economy*, 110, 37-72.
- [29] Diewert, W. E. (1973), Afriat and revealed preference theory, *Review of Economic Studies*, 40, 419-426.
- [30] Donni, O. (2007a), Household Behavior and Family Economics, *The Encyclopedia of Life Support Systems*, contribution 6.154.9.
- [31] Donni O. (2007b), Collective models of the household, *The New Palgrave, Dictionary of Economics, second edition*, forthcoming.
- [32] Duflo, E. (2003), Grandmothers and granddaughters: old age pension and intra-household allocation in South Africa, *World Bank Economic Review*, 17, 1-25.
- [33] Fortin, B. and G. Lacroix (1997), A test of the unitary and collective models of household labour supply, *Economic Journal*, 107, 933-955.
- [34] Grossbard, S. (2010), Independent individual decision-makers in household models and the New Home Economics, IZA working paper No. 5138.
- [35] Grossbard-Shechtman, A. (1984), A theory of allocation of time in markets for labor and marriage, *Economic Journal* 94, 863-882.
- [36] Grossbard-Shechtman, S. (2003), A consumer theory with competitive markets for work in marriage, *Journal of Socio-Economics* 31, 609-645.
- [37] Harbaugh, W.T., K. Krause and T. Berry (2001), GARP for Kids: On the development of rational choice behavior, *American Economic Review* 91, 1539-1545.
- [38] Lechene, V., Preston, I. (2005), Household Nash Equilibrium with Voluntary Contributed Public Goods, IFS working paper 05/06.

- [39] Lechene, V., Preston, I. (2010), Non Cooperative Household Demand, *Journal of Economic Theory*, forthcoming.
- [40] Lundberg, S. (1988), Labor supply of husbands and wives: a simultaneous equations approach, *Review of Economics and Statistics* 70, 224-235.
- [41] Lundberg, S. and R. Pollak (2007), Family Decision-Making, *The New Palgrave, Dictionary of Economics, second edition*, forthcoming.
- [42] McElroy, M. and M. Horney (1981), Nash-bargained household decisions: Towards a generalization of the theory of demand. *International Economic Review* 22, 333-349.
- [43] Samuelson, P. (1956) Social indifference curves", *Quarterly Journal of Economics*, 70, 1-22.
- [44] Selten, R (1991), Properties of a measure of predictive success, *Mathematical Social Sciences*, 153-167.
- [45] Sippel, R. (1997), An experiment on the pure theory of consumer behaviour, *Economic Journal* 107, 1431-44.
- [46] Thomas, D. (1990), Intra-household resource allocation: an inferential approach, *Journal of Human Resources*, 25, 635-664.
- [47] Varian, H.R. (1982), The nonparametric approach to demand analysis, *Econometrica* 50, 945-972.
- [48] Varian, H.R. (1990), Goodness-of-fit in optimizing models, *Journal of Econometrics* 46, 125-140.
- [49] Vermeulen, F. (2002), Collective household models: principles and main results, *Journal of Economic Surveys* 16, 533-564.

Table 1: experimental design - prices for the 9 choice problems

<i>Choice problem</i>	<i>Wine</i>	<i>Orange juice</i>	<i>M&Ms</i>
1	8	4	1
2	8	3	2
3	9	3	1
4	1	8	4
5	2	8	3
6	1	9	3
7	4	1	8
8	3	2	8
9	3	1	9

Notes: Prices are displayed in eurocents per commodity unit. A unit of red wine is 1 centiliter, a unit of orange juice is 3 centiliters, and a unit of M&Ms is 5 grams.

Table 2: experimental results - budget shares for the 9 choice problems

Choice	Wine	Orange juice	M&Ms	Wine	Orange juice	M&Ms	
		<i>102 individuals</i>			<i>51 dyads</i>		
1	<i>mean</i>	0.310	0.361	0.329	0.261	0.331	0.407
	<i>st. dev.</i>	0.296	0.248	0.297	0.291	0.226	0.282
2	<i>mean</i>	0.299	0.382	0.320	0.261	0.369	0.370
	<i>st. dev.</i>	0.292	0.260	0.292	0.303	0.245	0.283
3	<i>mean</i>	0.261	0.381	0.355	0.211	0.369	0.420
	<i>st. dev.</i>	0.289	0.276	0.305	0.271	0.247	0.299
4	<i>mean</i>	0.457	0.251	0.293	0.432	0.232	0.333
	<i>st. dev.</i>	0.354	0.255	0.301	0.360	0.253	0.315
5	<i>mean</i>	0.421	0.238	0.338	0.392	0.234	0.376
	<i>st. dev.</i>	0.323	0.234	0.293	0.329	0.233	0.299
6	<i>mean</i>	0.454	0.217	0.323	0.440	0.189	0.367
	<i>st. dev.</i>	0.346	0.250	0.308	0.367	0.241	0.328
7	<i>mean</i>	0.307	0.443	0.240	0.289	0.435	0.276
	<i>st. dev.</i>	0.315	0.321	0.301	0.303	0.304	0.315
8	<i>mean</i>	0.344	0.434	0.227	0.316	0.396	0.289
	<i>st. dev.</i>	0.303	0.294	0.292	0.308	0.261	0.305
9	<i>mean</i>	0.340	0.429	0.231	0.329	0.411	0.262
	<i>st. dev.</i>	0.318	0.321	0.316	0.332	0.332	0.342

Notes: For each choice problem, the table reports the mean budget shares (mean) over all participants (individuals and dyads), together with the corresponding standard deviations (st.dev.).

Table 3: experimental results - income sharing for the 9 choice problems

<i>Choice problem</i>	<i>Mean</i>	<i>St.dev.</i>	<i>Maximum</i>	<i>Equal split (percentage)</i>
1	0.126	0.156	0.680	29.41
2	0.119	0.143	0.513	31.37
3	0.105	0.146	0.552	33.33
4	0.135	0.195	0.800	35.29
5	0.090	0.122	0.506	37.25
6	0.106	0.154	0.532	31.37
7	0.076	0.089	0.360	39.22
8	0.079	0.101	0.460	35.29
9	0.079	0.106	0.410	45.10

Notes: For each choice problem, the table reports on the income share difference in relative terms (i.e. difference in income shares in absolute terms, divided by total income) over all dyads. It reports mean values, as well as corresponding standard deviations (st. dev.) and maximum values. In addition, the table gives the percentage of dyads that apply equal sharing (for each separate choice problem); see the column ‘equal split’.

Table 4: individual rationality - pass rate, power and predictive success

	<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
<i>Fit</i>	92.16	99.02	99.02
<i>Power</i>	82,00	65,86	59,69
Predictive success	74,16	64,88	58,71

Notes: 102 individuals in total.

Table 5: unitary rationality - pass rate, power and predictive success

	<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
<i>Fit</i>	90.20	98.04	100.00
<i>Power</i>	82.13	63.17	56.53
Predictive success	72.33	61.21	56.53

Notes: 51 dyads in total.

Table 6: collective rationality - pass rate, power and predictive success

		<i>a=0</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	62.75	92.16	98.04
	<i>Power</i>	97.42	93.16	88.43
	Predictive success	60.17	85.32	86.47
		<i>a=0.1</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	64.71	96.08	98.04
	<i>Power</i>	96.67	92.36	87.71
	Predictive success	61.38	88.44	85.75
		<i>a=0.2</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	70.59	96.08	98.04
	<i>Power</i>	96.05	91.65	86.87
	Predictive success	66.64	87.73	84.91
		<i>a=0.3</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	74.51	96.08	100.00
	<i>Power</i>	95.50	91.16	86.41
	Predictive success	70.01	87.24	86.41
		<i>a=0.4</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	78.43	96.08	100.00
	<i>Power</i>	95.18	90.69	85.93
	Predictive success	73.61	86.77	85.93
		<i>a=0.5</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	78.43	96.08	100.00
	<i>Power</i>	94.73	90.29	85.15
	Predictive success	73.16	86.37	85.15
		<i>a=1</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	82.35	98.04	100.00
	<i>Power</i>	93.68	89.06	83.60
	Predictive success	76.03	87.10	83.60
		<i>a=2</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	84.31	98.04	100.00
	<i>Power</i>	92.79	87.36	82.25
	Predictive success	77.10	85.40	82.25
		<i>a=4</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	84.31	98.04	100.00
	<i>Power</i>	92.15	86.49	81.38
	Predictive success	76.46	84.53	89.38
		<i>a=free</i>		
		<i>e=1</i>	<i>e=0.975</i>	<i>e=0.95</i>
	<i>Fit</i>	84.31	98.04	100.00
	<i>Power</i>	91.04	84.95	80.00
	Predictive success	75.35	82.99	80.00

Notes: 51 dyads in total.