Is utility transferable? A revealed preference analysis*

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

September 14, 2012

Abstract

The transferable utility hypothesis underlies important theoretical results in household economics. We provide a revealed preference framework for bringing this (theoretically appealing) hypothesis to observational data. First, we establish revealed preference conditions that must be satisfied for observed household consumption behavior to be consistent with transferable utility. Next, we show that these conditions are easily testable by means of integer programming methods. The test is entirely nonparametric, which makes it robust with respect to specification errors. Finally, we also provide a first empirical test of the transferable utility hypothesis through an application to Spanish household data.

JEL Classification: C14, DO1, D11, D12, D13.

Keywords: transferable utility hypothesis, generalized quasi-linearity, household consumption, nonparametric tests, revealed preferences.

1 Introduction

Household consumption analysis takes a prominent position in the microeconomics literature. In settings with multiple household members, theoretical consumption models often assume

*We thank Ian Crawford for generously providing us with the data of the Encuesta Continua de Presupestos Familiares (ECPF), which we use in our empirical study. We are also grateful to participants of the 2011 SAET conference in Ancao (Faro), the conference on "Bounded rationality, revealed preferences and consumer demand" in London and the "Annual Christmas meeting of Belgian economists" in Louvain-la-Neuve for useful discussion.

†Center for Economic Studies, University of Leuven. E. Sabbelaan 53, B-8500 Kortrijk, Belgium. email: laurens.cherchye@kuleuven.be. Laurens Cherchye gratefully acknowledges financial support from the Research Fund K.U.Leuven through the grant STRT1/08/004.

‡Center for Economic Studies, University of Leuven, E. Sabbelaan 53, B-8500 Kortrijk, Belgium. email: thomas.demuynck@kuleuven.be. Thomas Demuynck gratefully acknowledges the Fund for Scientific Research-Flanders (FWO-Vlaanderen) for his postdoctoral fellowship.

§ECARES-ECORE, Université Libre de Bruxelles. Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium. E-mail: bderock@ulb.ac.be. Bram De Rock gratefully acknowledges the European Research Council (ERC) for his Starting Grant.

transferable utility. As we will explain below, this assumption considerably simplifies the analysis. This paper provides a framework for bringing the (theoretically appealing) transferable utility hypothesis to empirical data. Specifically, we define the testable implications of transferable utility in revealed preference terms. In addition, we provide an application to observational household data, which gives a first empirical test of the transferable utility hypothesis.

The transferable utility hypothesis is a popular one in household economics. It underlies important theoretical results in the modeling of household behavior. Probably the best known example here is Becker (1974)'s Rotten Kid theorem; see Bergstrom (1989) for an insightful discussion. Bergstrom (1997) provides an extensive review of (other) applications of the transferable utility hypothesis in theoretical household models. Essentially, transferable utility means that it is possible to transfer utility from one household member to another member in a lossless manner, i.e. without affecting the aggregate household utility. Under transferable utility the frontier of the Pareto set is always a straight line of slope -1. This makes that the intrahousehold distribution of resources is independent of the aggregate household decisions: individual household members will always behave so as to maximize the size of the Pareto set.

The transferable utility assumption is popular because it has several highly desirable implications. First of all, it guarantees that household demand behavior displays attractive aggregation properties. In particular, any household then satisfies the so-called unitary model of household consumption, which means that aggregate household demand behaves as if it were generated by a single individual. However, as we will also discuss further on, consistency with the unitary model does not necessarily imply consistency with transferable utility, i.e. unitary household behavior is necessary but not sufficient for transferable utility. Next, the transferable utility hypothesis considerably facilitates welfare analysis. As the distribution of resources over the different household members does not influence the household decisions, welfare analysis can focus exclusively on the aggregate utility/welfare. Generally, utilizing the transferable utility hypothesis makes life of household economists a lot easier. Nevertheless, despite its wide prevalence in theoretical work, the empirical implications of transferable utility have hardly been studied. In fact, to the best of our knowledge, the hypothesis has never been tested on observational data.

This paper fills this gap: we develop tools for investigating the empirical realism of the transferable utility hypothesis. More specifically, we establish revealed preference conditions for observed consumption behavior to be consistent with the transferable utility assumption under Pareto efficient household behavior. These conditions are easily testable as they only require observations on consumed quantities at the household level and corresponding prices; testing the conditions can use standard integer programming methods. In addition, the test is entirely nonparametric, i.e. its empirical implementation does not require a prior (typically non–verifiable) functional structure for the utility functions of the individuals in the household.

We not only establish the testable conditions of transferable utility, but also provide a first empirical test of the hypothesis. Specifically, we apply our revealed preference conditions to Spanish household consumption data. We assess the empirical support for transferable utility by means of two complementary exercises. First, we conduct our basic revealed preference test for consistency of observed household behavior with transferable utility. This test is a 'sharp' one in that it abstracts from measurement error in the consumption data. As such, this first exercise will check whether observed consumption behavior is 'exactly' consistent with transferable utility. In our second exercise, we consider an extension of our basic test that does account for the possibility of measurement error. Here, we compute the minimal data adjustments that are needed for obtaining consistency with transferable utility. Essentially, for behavior that is not exactly consistent with transferable utility, this will tell us how 'close' it is to such consistency.

At this point, is worth indicating that the usefulness of our following results is not restricted to household settings. The transferable utility hypothesis is also pervasive in many other areas of economics. For example, in cooperative game theory the hypothesis is used to determine the value of a coalition and underlies notions such as the Shapley value (Shapley, 1953), the kernel (Davis and Maschler, 1965) and the nucleolus (Schmeidler, 1969). Next, it is a critical assumption in the Shapley-Shubik assignment model (Shapley and Shubik, 1972), which has become a workhorse in the study of labor and marriage markets and other models of two–sided matching. Furthermore, transferable utility is crucial for the validity of the famous Coase theorem (see Coase (1960) and Hurwicz (1995)). Lastly, the hypothesis is also frequently used in principal–agent models, theoretical mechanism design, matching models, public economics, industrial organization, international trade, and so on.

The remainder of the paper unfolds as follows. In Section 2, we briefly recapture some important building blocks for our following analysis, and we articulate our own contributions to the existing literature. Here, we will also indicate that the so-called generalized quasi-linear (GQL) utility specification provides a necessary and sufficient condition for a Pareto optimal household allocation rule to be consistent with transferable utility. In Section 3, we then formally define this GQL specification. Section 4 subsequently presents the corresponding revealed preference characterization. Section 5 provides the integer programming formulation of our characterization and presents the empirical application. Finally, Section 6 concludes.

2 Testable implications of transferable utility

Generalized quasi-linearity. To define the testable implications of transferable utility at the household level, we need to characterize the underlying utility functions of the individuals within the household. The best–known specification leading to the property of transferable utility is the quasi-linear (QL) utility specification. This specification requires the utility functions of the individuals to be linear in at least one good, usually called the numeraire. Unfortunately, QL utility has strong and unrealistic implications (e.g. absence of income effects for all but a single good, risk neutrality, etc.).

In the presence of public goods, Bergstrom and Cornes (1981, 1983) and Bergstrom (1989) showed that a weaker form than QL utility equally implies transferable utility, i.e. so-called 'generalized' quasi-linear (GQL) utility (a term coined by Chiappori (2010)). Interestingly, these authors also showed that this GQL specification provides a necessary and sufficient condition for transferable utility under Pareto efficient household behavior. The GQL form can be obtained from the QL specification through multiplication of the numeraire by a function defined in terms of the bundle of (intra-household) public goods. The additional requirement that this function is common to all individuals within the household provides the property of transferable utility. As households typically consume a large amount of public goods, this characterization of transferable utility is particularly convenient in household settings.

Recently, Chiappori (2010) derived a set of necessary and sufficient conditions on the (aggregate) household demand function such that it is compatible with a Pareto efficient allocation where household members are endowed with GQL utility functions. As far as we know, this is the first (and –up till now– sole) study that makes the testable implications of transferable utility explicit. In view of our following exposition, we remark that Chiappori adopted a so-called 'differential' approach to characterizing GQL utility: he focused on testable (differential) properties of the household demand function to be consistent with transferable utility. Practical applications of this differential approach then typically require a prior parametric specification of this demand function, which is to be estimated from the data. As we will indicate below, this implies a most notable difference with the approach that we follow here.

Revealed preference implications. We complement Chiappori's findings by establishing testable conditions of transferable utility (or GQL utility) in the revealed preference tradition of Samuelson (1938), Houthakker (1950), Afriat (1967), Diewert (1973) and Varian (1982). In contrast to the differential approach, this revealed preference approach obtains conditions that can be verified by (only) using a finite set of household consumption observations (i.e. prices and quantities) and, thus, it does not require the estimation of a household demand function. As such, a main advantage of these revealed preference conditions is that they allow a nonparametric analysis of the data: they do not impose any functional form on the utility function (generating a particular household demand function) except for usual regularity conditions.

More specifically, we get necessary and sufficient conditions that enable checking consistency of a given data set with transferable utility. In the spirit of Varian (1982), we refer to this as 'testing' data consistency with transferable utility. As for the practical application of this test, we will show that our revealed preference conditions can be equivalently reformulated as integer programming constraints. This integer programming formulation allows us to test data consistency with transferable utility by applying standard integer programming solution techniques.

Our empirical application in Section 5 will demonstrate the empirical usefulness of this integer programming approach. In this respect, it is worth pointing out that a similar integer programming approach has been fruitfully applied for a revealed preference analysis of household consumption behavior; see, for example, Cherchye, De Rock, and Vermeulen (2011) and Cherchye, Demuynck, and De Rock (2010). Our study complements these earlier studies by further illustrating the versatility of integer programming techniques for analyzing household consumption in revealed preference terms. Cherchye, Crawford, De Rock, and Vermeulen (2009a) also provide a theoretical analysis of the computational complexity of revealed preference conditions for general equilibrium models that are formally close to the transferable utility conditions that we study here. In principle, this analysis could be translated to our setting but, for compactness, we will not do this in the current paper.

Further contributions. At this point, it is worth to indicate two further important differences between our study and the original study of Chiappori (2010), which involve two additional contributions. First, to establish his characterization, Chiappori assumed observability of the numeraire good. However, in practice this numeraire good is typically an 'outside' good, i.e. the amount of money not spent on observed consumption, which is usually not recorded in real–life applications (including our own application). Given this, our following revealed preference analysis will principally focus on characterizing transferable utility for the case with an unobserved numeraire (or outside good). To obtain this characterization, we will first have to establish the characterization that applies to an observed numeraire.

Another main difference between our study and the one of Chiappori is that we present an empirical application that effectively brings our testable implications to observational data. As indicated above, as far as we know, this provides a first empirical test of the transferable utility hypothesis. Specifically, we verify our revealed preference conditions for a sample of households drawn from the Encuesta Continua De Presupestos Familiares (ECPF), a Spanish consumer expenditure survey. In general, our results are mixed. Although we find the assumption of transferable utility to be realistic for a considerable part of the households under consideration, there is also a substantial share of households whose behavior contradicts the transferable utility assumption. Given this, we will explore the possibility to rationalize observed consumption behavior by accounting for (a little) measurement error.

As a final remark, we indicate that Brown and Calsamiglia (2007) recently developed a revealed preference characterization of the QL utility specification. By focusing on the GQL utility form, we provide revealed preference conditions for a model that contains this QL specification as a special case. In addition, in our empirical exercise we will compare the empirical goodness of the GQL and QL specifications. A main conclusion here will be that, for the given households, the GQL utility specification provides a better fit of the observed consumption behavior than the QL specification.

3 Generalized quasi-linear utility

Consider a household with M (≥ 2) members. Each member m ($\leq M$) consumes a bundle of $N+1$ private goods $(\mathbf{q}^m, x^m) \in \mathbb{R}_+^{N+1}$ and a bundle of K public goods $\mathbf{Q} \in \mathbb{R}_+^K$. The private good x^m denotes member m's amount of the numeraire. For each m, we assume $x^m > 0$ in what follows.[1] In addition, we normalize by setting the price of the numeraire equal to one. Next, the vector $\mathbf{p} \in \mathbb{R}_{++}^N$ represents the normalized price vector for the bundle of private goods $\mathbf{q}^m$, while the vector $\mathbf{P} \in \mathbb{R}_{++}^K$ is the normalized price vector for the bundle of public goods $\mathbf{Q}$.

Utility of member m is represented by the strictly increasing and quasi–concave utility function $u^m(\mathbf{q}^m, x^m, \mathbf{Q})$. The utility functions u^m are said to be of the generalized quasi–linear (GQL) form if there exist a (member-specific) function $b^m : \mathbb{R}_+^{K+N} \to \mathbb{R}$ and a (common) function $a : \mathbb{R}_+^K \to \mathbb{R}_{++}$ such that

$$u^m(\mathbf{q}^m, x^m, \mathbf{Q}) = a(\mathbf{Q})x^m + b^m(\mathbf{Q}, \mathbf{q}^m). \quad (1)$$

Bergstrom and Cornes (1983) have shown that member-specific GQL utilities are necessary

[1] Just like for quasi-linearity, we need a non-zero amount of the numeraire in order to have transferable utility.

and sufficient for transferable utility under Pareto efficient household behavior.[2]

The GQL specification encompasses the quasi-linear (QL) specification as a special case. Specifically, if $a(\mathbf{Q}) = a$ for all $\mathbf{Q}$ (i.e. the function value $a(\mathbf{Q})$ is everywhere the same) then the specification in (1) coincides with the QL specification:

$$u^m(\mathbf{q}^m, x^m, \mathbf{Q}) = a\ x^m + b^m(\mathbf{Q}, \mathbf{q}^m).$$

However, if $a(\mathbf{Q})$ varies with the level of public goods, then the GQL specification vastly expands the range of utility functions compatible with transferable utility.

We assume that household decisions are made according to the Pareto criterion: allocations are chosen such that no member can be made better of without reducing the utility of some other household member.[3] In this case, any equilibrium allocation $(\mathbf{q}^1, \ldots, \mathbf{q}^M, x^1, \ldots, x^M, \mathbf{Q})$ minimizes total household expenditures subject to the constraint that every member of the household receives at least some predefined level of utility $\bar{u}^m$. In other words, given a fixed vector of utility levels $(\bar{u}^1, \ldots, \bar{u}^M) \in \mathbb{R}^M_+$, Pareto efficiency imposes that the household decision making process solves the next optimization problem (**OP.1**):

$$\min_{(\mathbf{q}^1,\ldots,\mathbf{q}^M,x^1,\ldots,x^M,\mathbf{Q}) \in \mathbb{R}_+^{M(N+1)+K}} \sum_{m=1}^{M} x^m + \sum_{m=1}^{M} \mathbf{p}\mathbf{q}^m + \mathbf{P}\mathbf{Q}$$
$$\text{s.t. } a(\mathbf{Q})x^m + b^m(\mathbf{Q}, \mathbf{q}^m) \geq \bar{u}^m \ (\forall m \leq M).$$

In view of our following analysis, we develop an equivalent formulation of **OP.1**. To obtain the formulation, we first observe that each constraint will be binding in the solution of **OP.1** because the utility functions u^m are strictly increasing. Using this, and because $x^m > 0$ for all m, we can substitute the restrictions in the objective function. As a result, we can equivalently reformulate the original optimization problem as follows (**OP.2**):

$$\min_{(\mathbf{q}^1,\ldots,\mathbf{q}^M,\mathbf{Q}) \in \mathbb{R}_+^{MN+K}} \alpha(\mathbf{Q}) \sum_{m=1}^{M} \bar{u}^m - \sum_{m=1}^{M} \beta^m(\mathbf{q}^m, \mathbf{Q}) + \sum_{m=1}^{M} \mathbf{p}\mathbf{q}^m + \mathbf{P}\mathbf{Q},$$
$$\text{with } \alpha(\mathbf{Q}) = \frac{1}{a\left(\mathbf{Q}\right)} \text{ and } \beta^m(\mathbf{q}^m, \mathbf{Q}) = \frac{b^m(\mathbf{q}^m, \mathbf{Q})}{a\left(\mathbf{Q}\right)} \ \ (\forall m \leq M)\,.$$

From this equivalent formulation, it is directly clear that the optimal solution of problem **OP.1** only depends on the total amount of utility $\sum_m^M \bar{u}^m$ but not on the specific distribution of this amount over the different household members. This demonstrates the property of transferable utility under GQL.

Standard first order conditions characterize the (interior) solutions of problem **OP.2** if the function α is convex and the functions β^m are concave. Bergstrom and Cornes (1983) showed that these requirements are equivalent to the condition that the utility functions u^m are quasi-concave (which we assumed before). Next, it is easy to verify that α is decreasing in

[2]See also Browning, Chiappori, and Weiss (2011, p. 276) for a detailed discussion of this functional specification.

[3]See Chiappori (1988) and Cherchye, De Rock, and Vermeulen (2007, 2009b, 2011) for revealed preference tests of the assumption of Pareto optimality, without the additional assumption of transferable utility.

$\mathbf{Q}$ while the β^m are increasing in $\mathbf{q}$. If we further assume that b^m and a are bounded from below and a is strictly positive, then β^m is also increasing in $\mathbf{Q}$.[4] For an optimal solution $(\mathbf{q}^{1*}, \dots, \mathbf{q}^{M*}, x^{1*}, \dots, x^{M*}, \mathbf{Q}^*)$ of problem **OP.2**, the first order conditions are as follows:[5]

$$-\frac{\partial \alpha(\mathbf{Q}^*)}{\partial \mathbf{Q}} \sum_{m=1}^{M} \bar{u}^m + \sum_{m=1}^{M} \frac{\partial \beta^m(\mathbf{q}^{m*}, \mathbf{Q}^*)}{\partial \mathbf{Q}} = \mathbf{P}, \tag{foc.1}$$

$$\frac{\partial \beta^m(\mathbf{q}^{m*}, \mathbf{Q}^*)}{\partial \mathbf{q}^m} = \mathbf{p}, \tag{foc.2}$$

$$\frac{x^{m*}}{\alpha(\mathbf{Q}^*)} + \frac{\beta(\mathbf{q}^{m*}, \mathbf{Q}^*)}{\alpha(\mathbf{Q}^*)} = \bar{u}^m. \tag{foc.3}$$

Conditions (foc.1) and (foc.2) provide a formal expression of the household's marginal decision rules for the public and private goods, respectively. Next, condition (foc.3) complies with the GQL utility specification in (1). The first order conditions (foc.1)–(foc.3) provide a useful starting point for developing our revealed preference characterization in the next section.

4 Revealed preference characterization

We analyze the (aggregate) consumption behavior of a household with M individuals, by starting from a finite set T of observed household choices. For each observation $t \in T$, we know the privately and publicly consumed quantities $\mathbf{q}_t$ and $\mathbf{Q}_t$, as well as the corresponding prices $\mathbf{p}_t$ and $\mathbf{P}_t$. Remark that we only observe the aggregate private quantities $\mathbf{q}_t$ and not the member-specific quantities $\mathbf{q}_t^m$. In a first instance we assume that the aggregate amount of the numeraire ('outside') good at every t (i.e. x_t) is also observed (again we assume that the member-specific quantities x_t^m are not observed). We will relax this assumption later on. As discussed before, we believe an unobserved numeraire is a more realistic assumption for real life applications.

Numeraire observed. If the consumption of the numeraire is observed, then the relevant data set is $S = \{\mathbf{p}_t, \mathbf{P}_t; x_t, \mathbf{q}_t, \mathbf{Q}_t\}_{t \in T}$. In what follows, we present necessary and sufficient conditions for the set S to be rationalizable in terms of GQL utility functions, i.e. there exist functions α and β^m so that each bundle $(x_t, \mathbf{q}_t, \mathbf{Q}_t)$ $(t \in T)$ leads to a solution for **OP.2**. This provides a characterization of transferable utility in the revealed preference tradition. Our starting definition is the following:

Definition 1 (TU-rationalizable) *The data set $S = \{\mathbf{p}_t, \mathbf{P}_t; x_t, \mathbf{q}_t, \mathbf{Q}_t\}_{t \in T}$ is transferable utility (TU)-rationalizable if (i) there exist a convex and decreasing function $\alpha : \mathbb{R}_+^K \to \mathbb{R}$ and*

[4] We can show this by contradiction. Assume that β^m is non-increasing in $\mathbf{Q}$ at some bundle. Then, concavity of β^m implies that β^m is unbounded from below. However, as a is strictly positive for all $\mathbf{Q}$, this means that b^m must be unbounded from below, which gives the wanted contradiction. We thank Phil Reny for pointing this out to us.

[5] If α or β are not differentiable we may take the sub- and superdifferentials that satisfy the corresponding first order conditions. We will also use this in the proof of Proposition 1.

M concave and increasing functions $\beta^m : \mathbb{R}_+^{N+K} \to \mathbb{R}^N$ and (ii), for each t, there exist private consumption bundles $\mathbf{q}_t^1, \ldots, \mathbf{q}_t^M$ that sum to $\mathbf{q}_t$ and strictly positive numbers $x_t^1, \ldots, x_t^M$ that sum to x_t such that $\{\mathbf{q}_t^1, \ldots, \mathbf{q}_t^M, \mathbf{Q}_t\}$ solves OP.2 given the prices $\mathbf{p}_t, \mathbf{P}_t$ and utility levels $\bar{u}_t^m = \dfrac{x_t^m}{\alpha(\mathbf{Q}_t)} + \dfrac{\beta^m(\mathbf{q}_t^m, \mathbf{Q}_t)}{\alpha(\mathbf{Q}_t)}$.

Of course, the above definition could equally well have been stated by using the functions a and b^m and by referring to program **OP.1**. We opt for the current statement to enhance the interpretation of the revealed preference characterization below.

It follows from Definition 1 that the concept of TU-rationalizability implicitly depends on the number of individuals within the household. However, as the following result shows, this qualification is actually irrelevant in view of practical applications: it is empirically impossible to distinguish between different household sizes; there exists a rationalization of the set S in terms of a single individual (i.e. $M = 1$) if and only if there exists one in terms of any number of individuals. More specifically, we can prove the following result:[6]

Proposition 1 *Consider a data set $S = \{\mathbf{p}_t, \mathbf{P}_t; x_t, \mathbf{q}_t, \mathbf{Q}_t\}_{t \in T}$. The following statements are equivalent:*

1. *The data set S is TU-rationalizable for a household of M individuals;*

2. *The data set S is TU-rationalizable for a household of a single individual;*

3. *For all $t \in T$, there exists $\alpha_t \in \mathbb{R}_{++}$, $\beta_t, \bar{u}_t \in \mathbb{R}_+$, $\boldsymbol{\lambda}_t^\alpha \in \mathbb{R}_-^K$ and $\boldsymbol{\lambda}_t^\beta \in \mathbb{R}_{++}^K$ such that, for all $t, v \in T$:*

$$\alpha_t - \alpha_v \geq \boldsymbol{\lambda}_v^\alpha(\mathbf{Q}_t - \mathbf{Q}_v), \tag{RP.1}$$

$$\beta_t - \beta_v \leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + \boldsymbol{\lambda}_v^\beta(\mathbf{Q}_t - \mathbf{Q}_v), \tag{RP.2}$$

$$\boldsymbol{\lambda}_t^\beta - \boldsymbol{\lambda}_t^\alpha \bar{u}_t = \mathbf{P}_t, \tag{RP.3}$$

$$\bar{u}_t = \frac{x_t}{\alpha_t} + \frac{\beta_t}{\alpha_t}. \tag{RP.4}$$

The equivalence between statements 1 and 2 demonstrates the aggregation property of the transferable utility assumption that we mentioned above: if a data set is TU-rationalizable for a household of M individuals, then it is rationalizable for a single individual (endowed with a GQL utility function), and vice versa.[7] Statement 3 then provides the combinatorial conditions that characterize the collection of data sets that are TU-rationalizable. The first two conditions ((RP.1) and (RP.2)) define so-called Afriat inequalities (see also our discussion of Afriat's Theorem in Appendix C) that apply to our specific setting. In terms of Definition 1 these inequalities correspond to, respectively, the (convex) function α and the (concave) function β (where we drop the index m because of the equivalence between statements 1 and 2). The vectors $\boldsymbol{\lambda}_t^\alpha$ and $\boldsymbol{\lambda}_t^\beta$ then represent the gradient vectors of these functions in terms of the public goods bundle. Finally, the conditions (RP.3) and (RP.4) give the revealed preference counterparts of the first order conditions (foc.1) and (foc.3) that we discussed in the previous section.

[6] Appendix A contains the proofs of our main results.
[7] Chiappori (2010) obtained a similar result in his differential setting.

Numeraire unobserved. In real life applications the amount of the numeraire good is usually not observed. For example, this will also be the case in our own application. The relevant data set is then given as $S = \{\mathbf{p}_t, \mathbf{P}_t; \mathbf{q}_t, \mathbf{Q}_t\}_{t\in T}$.

Interestingly, the result in Proposition 1 enables us to establish a characterization of transferable utility for such a data set S. Specifically, we can derive the following result:

Proposition 2 *Consider a data set $S = \{\mathbf{p}_t, \mathbf{P}_t; \mathbf{q}_t, \mathbf{Q}_t\}_{t\in T}$. The following statements are equivalent:*

1. *For all $t \in T$, there exist $x_t \in \mathbb{R}_{++}$ such that $\{\mathbf{p}_t, \mathbf{P}_t; x_t, \mathbf{q}_t, \mathbf{Q}_t\}_{t\in T}$ is TU-rationalizable for a household of M individuals (or, equivalently, a single individual);*

2. *For all $t \in T$, there exist $U_t^A, U_t^B \in \mathbb{R}_+$, $\lambda_t^A \in \mathbb{R}_{++}$, $\mathbf{P}_t^A \in \mathbb{R}_+^K$, $\mathbf{P}_t^B \in \mathbb{R}_{++}^K$ such that, for all $t, v \in T$:*

$$U_t^A - U_v^A \leq \lambda_t^A \left[\mathbf{P}_v^A(\mathbf{Q}_t - \mathbf{Q}_v)\right], \tag{RP.5}$$
$$U_t^B - U_v^B \leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + \mathbf{P}_v^B(\mathbf{Q}_t - \mathbf{Q}_v), \tag{RP.6}$$
$$\mathbf{P}_t^A + \mathbf{P}_t^B = \mathbf{P}_t. \tag{RP.7}$$

When compared to the characterization in Proposition 1, the conditions (RP.5), (RP.6) and (RP.7) in Proposition 2 correspond to (RP.1), (RP.2) and (RP.3), respectively. We refer to the proof of the result for an explicit construction. This proof also shows that, for each observation t, we can always construct a numeraire quantity x_t that meets condition (RP.4) if the data satisfy (RP.5)–(RP.7).

Nested models. To conclude this section, we discuss the relationship between the transferable utility conditions developed above and closely related rationalizability conditions that have been considered in the revealed preference literature. Specifically, we make explicit how the transferable utility model is situated 'between' the quasi-linear (QL) utility model and the unitary model. This further clarifies the interpretation of our revealed preference characterization of transferable utility.

As a first exercise, we recall from the previous section that QL utility imposes that the function value $\alpha(\mathbf{Q})$ is constant for all $\mathbf{Q}$. In terms of the characterization in Proposition 1, this means that the gradient vector $\boldsymbol{\lambda}_t^\alpha$ equals zero. One can then easily verify that the conditions (RP.1)-(RP.4) reduce to

$$\beta_t - \beta_v \leq \mathbf{p}_t(\mathbf{q}_t - \mathbf{q}_v) + \mathbf{P}_t(\mathbf{Q}_t - \mathbf{Q}_v). \tag{RP.8}$$

This condition is necessary and sufficient for data consistency with the QL utility specification.[8] We observe that the QL condition (RP.8) is independent of the level of the numeraire (x_t), which implies a notable difference with our above characterization of GQL utility. In fact, this independence is also revealed by the fact that the conditions (RP.5)–(RP.7) in Proposition 2

[8] In fact, condition (RP.8) is equivalent to the revealed preference condition that Brown and Calsamiglia (2007) originally derived for data consistency with the QL specification.

equally coincide with (RP.8) if we set $\mathbf{P}_t^A$ equal to zero for all $t \in T$ (which has a similar meaning as $\boldsymbol{\lambda}_t^\alpha = \mathbf{0}$ in Proposition 1).

Next, it directly follows from statement 2 in Proposition 1 that the transferable utility model is nested in the unitary model. In fact, in Appendix B we show that conditions (RP.1)-(RP.4) automatically require that the data satisfy the Generalized Axiom of Revealed Preference (GARP), which is necessary and sufficient for data consistency with the unitary model.[9] In other words, if a household data set is TU-rationalizable then the household acts as a single individual. However, a household may well behave as if it were a single decision maker without satisfying transferable utility. In this sense, our revealed preference conditions in Propositions 1 and 2 capture the *additional* restrictions that observed consumption behavior must satisfy for the transferable utility assumption to hold. Our conditions effectively allow for bringing these specific restrictions of TU-rationalizability to empirical data.

In this respect, one further point relates to Proposition 2. This result makes clear that transferable utility has testable implications even if the numeraire good is not observed. (And, as we will show in the next section, these implications may actually be fairly strong for observational data.) By contrast, following a revealed preference approach similar to ours, Varian (1988) has shown that the unitary model does not have any testable implications as soon as we do not observe the consumption quantity of some good (in casu the numeraire quantity x_t). We believe this is an interesting observation, as it suggests that considering the transferable utility model may be empirically meaningful even if the unitary model is non-testable. (In fact, it also motivates why we will not consider tests of the unitary model in our following application, which focuses on a setting with an unobserved numeraire.)

5 Empirical application

In this section, we will use the above revealed preference characterization to empirically assess the validity of the transferable utility hypothesis for Spanish household data. As indicated in the Introduction, our application consists of two complementary exercises. In a first step, we consider the basic revealed preference tests as they have been introduced above. It follows from our exposition that these tests are 'sharp', in the following sense: either a data set exactly satisfies the data consistency conditions or it does not; the tests abstract from possible measurement error in the consumption data. In this exercise, an important focus will be on comparing the empirical performance of the GQL and QL utility specifications. Here, a main result will be that a considerable part of the households under evaluation passes the GQL (or transferable utility) test while failing the QL test. To fairly assess this difference, we also consider the discriminatory power of the two tests under study; this should account for a possible trade–off between power and pass rates.

In our second step, we then introduce an extension of our basic tests that does account for the possibility of measurement error. Specifically, we compute the minimal data adjustments

[9] See the next section (Definition 2) for a formal definition of GARP. In this respect, we also refer to Afriat's Theorem that we recapture in Appendix C. This result states the conditions for a price-quantity set to be rationalizable in terms of a single utility function. As such, it actually provides the revealed preference characterization of the unitary model that complements our characterization of the transferable utility model.

that we need for making behavior consistent with transferable utility (i.e. rationalizable in terms of GQL utility functions). Essentially, if the observed behavior of some household does not exactly fit transferable utility, this will tell us how 'close' the household data are to such an exact fit. Like before, we do not merely study the GQL specification, but also compare the results for this specification with those for the QL specification. This should provide complementary information on the empirical validity of transferable utility as well as the relative empirical performance of the two utility specifications.

As we motivated before, our following analysis will concentrate on the case where the quantity of the numeraire good is not observed. Before presenting our data and results, we first introduce the integer programming formulation of the conditions (RP.5)–(RP.7) in Proposition 2.

Integer programming formulation. The conditions (RP.6) and (RP.7) in Proposition 2 are linear and therefore easily verifiable, while the Afriat inequalities in condition (RP.5) are quadratic (i.e. nonlinear in the unknown λ_t's and $\mathbf{P}_t^A$'s). From a practical point of view, this nonlinearity makes it difficult to empirically verify the characterization in Proposition 2. However, these Afriat inequalities can be equivalently restated in terms of linear (mixed) binary integer programming constraints by making use of the Generalized Axiom of Revealed Preferences (GARP); this follows from Afriat's Theorem that we recapture in Appendix C. As is well known, linear (mixed) binary integer programming problems can be solved more efficiently than programs with quadratic constraints.[10]

Let us consider a general setting with a set $Z = \{\mathbf{w}_l; \mathbf{y}_l\}_{l \in L}$ containing (strictly positive) price vectors $\mathbf{w}_l$ and (positive) quantity vectors $\mathbf{y}_l$. Then the GARP condition is as follows:

Definition 2 *Consider a set $Z = \{\mathbf{w}_l; \mathbf{y}_l\}_{l \in L}$. For any $l_1, l_2 \in L$, $\mathbf{y}_{l_1} R \mathbf{y}_{l_2}$ if $\mathbf{w}_{l_1}\mathbf{y}_{l_1} \geq \mathbf{w}_{l_1}\mathbf{y}_{l_2}$. Next, $\mathbf{y}_{l_1} R \mathbf{y}_{l_2}$ if there exists a sequence $r, \ldots, t$ (with $r, \ldots, t \in L$) such that $\mathbf{y}_{l_1} R \mathbf{y}_r$,...,$\mathbf{y}_t R \mathbf{y}_{l_2}$. The set Z satisfies GARP if, for all $l_1, l_2 \in L$, $\mathbf{y}_{l_1} R \mathbf{y}_{l_2}$ implies $\mathbf{w}_{l_2}\mathbf{y}_{l_1} \geq \mathbf{w}_{l_2}\mathbf{y}_{l_2}$. We refer to R as a revealed preference relation.*

We now have the following proposition, which makes use of the binary variables $r_{t,v}$.

Proposition 3 *Consider a data set $S = \{\mathbf{p}_t, \mathbf{P}_t; \mathbf{q}_t, \mathbf{Q}_t\}_{t \in T}$. The following statements are equivalent:*

1. *For all $t \in T$, there exist $x_t \in \mathbb{R}_{++}$ such that $\{\mathbf{p}_t, \mathbf{P}_t; x_t, \mathbf{q}_t, \mathbf{Q}_t\}_{t \in T}$ is TU-rationalizable for a household of M individuals (or, equivalently, a single individual);*

2. *For all $t, v \in T$, there exist $r_{t,v} \in \{0, 1\}, U_t^A, U_t^B \in \mathbb{R}_+$, $\mathbf{P}_t^A \in \mathbb{R}_+^K$, $\mathbf{P}_t^B \in \mathbb{R}_{++}^K$ such that,*

[10] Specifically, by adding for any binary variable r the constraints $0 \leq r \leq 1$ and $r^2 = r$, we can easily convert any linear (mixed) binary integer programming problem into a problem with quadratic constraints.

for all $t, v, s \in T$:

$$U_t^B - U_v^B \leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + \mathbf{P}_v^B(\mathbf{Q}_t - \mathbf{Q}_v), \tag{IP.1}$$

$$\mathbf{P}_t^A + \mathbf{P}_t^B = \mathbf{P}_t, \tag{IP.2}$$

$$\mathbf{P}_t^A(\mathbf{Q}_t - \mathbf{Q}_v) < r_{t,v}C, \tag{IP.3}$$

$$r_{t,v} + r_{v,s} \leq 1 + r_{t,s}, \tag{IP.4}$$

$$\mathbf{P}_t^A(\mathbf{Q}_t - \mathbf{Q}_v) \leq (1 - r_{v,t})C, \tag{IP.5}$$

with C *a given number exceeding all observed* $\mathbf{P}_t\mathbf{Q}_t$.

The linear inequalities (IP.1) and (IP.2) are clearly identical to (RP.6) and (RP.7). Further, the nonlinear inequalities (RP.5) have been replaced by the linear inequalities (IP.3)–(IP.5) that make use of real and binary variables. More specifically, (IP.3)-(IP.5) correspond to the GARP condition in Definition 2 in which we take $\mathbf{w}_l = \mathbf{P}_l^A$ and $\mathbf{y}_l = \mathbf{Q}_l$.

To explain the inequalities (IP.3)-(IP.5), we interpret the variables $r_{t,v}$ in terms of the revealed preference relation R, i.e. $r_{t,v} = 1$ corresponds to $\mathbf{Q}_t R \mathbf{Q}_v$. The constraint (IP.3) then imposes $\mathbf{Q}_t R \mathbf{Q}_v$ (or $r_{t,v} = 1$) whenever $\mathbf{P}_t^A\mathbf{Q}_t \geq \mathbf{P}_t^A\mathbf{Q}_v$. Next, the constraint (IP.4) complies with transitivity of the relation R: if $\mathbf{Q}_t R \mathbf{Q}_v$ ($r_{t,v} = 1$) and $\mathbf{Q}_v R \mathbf{Q}_s$ ($r_{v,s} = 1$), then $\mathbf{Q}_t R \mathbf{Q}_s$ ($r_{t,s} = 1$). Finally, the constraint (IP.5) states that, if $\mathbf{Q}_v R \mathbf{Q}_t$ ($r_{v,t} = 1$), then we must have $\mathbf{P}_t^A\mathbf{Q}_t \leq \mathbf{P}_t^A\mathbf{Q}_v$.

For a given data set S, we can verify the above linear inequalities by using mixed integer linear programming techniques. Given the result in Proposition 3, this effectively checks whether the set S is consistent with transferable utility (i.e. rationalizable in terms of GQL utility functions).

Application set-up. Our data are drawn from the Encuesta Continua de Presupestos Familiares (ECPF). The ECPF is a quarterly budget survey (1985–1997) that interviews about 3200 Spanish households on their consumption expenditures. For each household, the data provides consumption observations for a maximum of eight consecutive quarters. See Browning and Collado (2001) and Crawford (2010) for a more detailed explanation of this data set.

For obvious reasons, we focus on households with at least two household members. Next, all households in our sample are headed by a married couple where the husband is full time employed and the wife is outside the labor force. Finally, we exclude all households with less than eight observations. In the end, this obtains a panel with 1585 households.

For each household, we have consumption data (quantities and prices) for 15 nondurable consumption goods: (i) food and non-alcoholic drinks at home; (ii) alcohol; (iii) tobacco; (iv) energy at home (heating by electricity); (v) services at home (heating: not electricity, water, furniture repair); (vi) nondurables at home (cleaning products); (vii) non-durable medicines; (viii) medical services; (ix) transportation; (x) petrol; (xi) leisure (cinema, theater, clubs for sports); (xii) personal services; (xiii) personal nondurables (toothpaste, soap); (xiv) restaurants and bars and (xv) traveling (holiday). We will treat energy at home, services at home and nondurables at home as our three public goods. To obtain normalized prices, we deflate the price (index) for each good (category) by the value of the consumer price index in the corresponding quarter.

To avoid (debatable) preference homogeneity assumptions across similar households, we will consider each household separately in our following analysis. In other words, we consider a different data set $S = \{\mathbf{p}_t, \mathbf{P}_t; \mathbf{q}_t, \mathbf{Q}_t\}_{t \in T}$ for every individual household. This practice effectively accounts for inter–household heterogeneity and, thus, optimally exploits the panel structure of our data set.

Pass rates and power. We first consider the basic tests of our rationalizability conditions, which do not account for measurement error in the household consumption data. In what follows, we do not only consider the mere test results but also the discriminatory power of the two (GQL and QL) rationalizability conditions under study. Indeed, Bronars (1987) and, more recently, Andreoni and Harbaugh (2008) and Beatty and Crawford (2011) –rather convincingly– argue that revealed preference test results (indicating pass or fail of the data for some behavioral condition) should be complemented with power measures to obtain a fair empirical assessment of the condition under evaluation. Favorable test results (i.e. a high pass rate for some given data), which prima facie suggest a good empirical fit, have little value if the test has little discriminatory power (i.e. the condition is hard to reject for the data at hand).

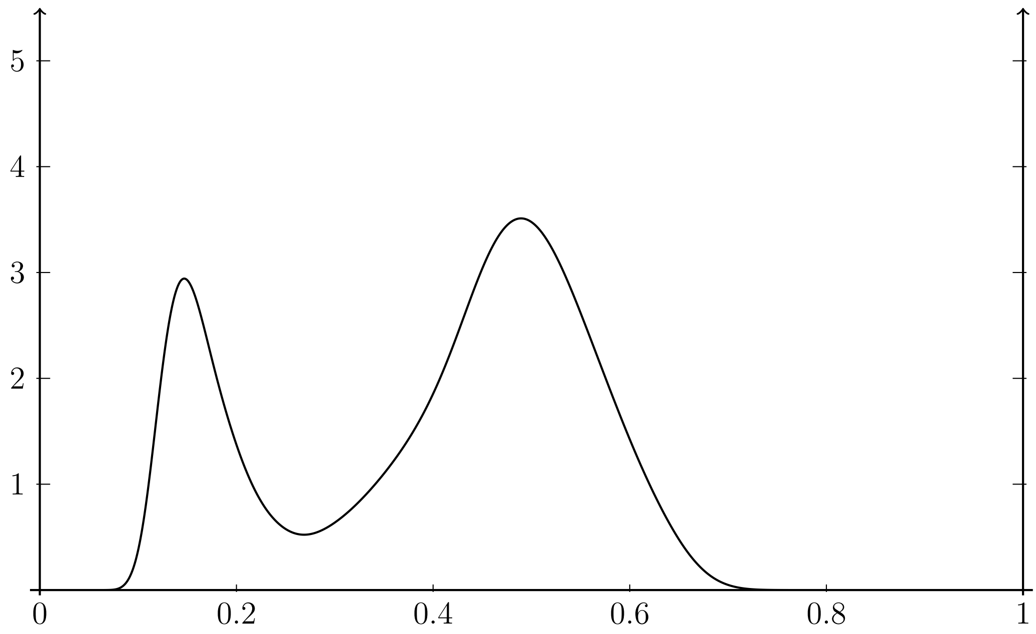
For both rationalizability conditions under evaluation, we compute a power measure for every individual household. This measure quantifies discriminatory power in terms of the probability to detect random behavior, and is constructed as follows. We model random behavior by using a bootstrap procedure: we simulate 1000 random series of eight consumption choices by constructing, for each of the eight observed household budgets, a random quantity bundle exhausting the given budget (for the corresponding prices). We construct these quantity bundles by randomly drawing (with replacement) budget shares (for the 15 goods) from the set of 12680 (= 8 x 1585) observed household choices in our data set. The power measure is then calculated as one minus the proportion of these randomly generated consumption series that are consistent with the rationalizability condition under evaluation. By using this bootstrap method, our power assessment gives information on the expected distribution of violations under random choice, while incorporating information on the households' actual choices.[11]

Table 1 presents our results. The first column in the table gives the pass rates for the two rationalizability conditions. Each pass rate gives the proportion of households in our sample that meets a particular (GQL or QL) condition. A first observation is that the GQL condition does not allow us to rationalize all consumption behavior in our sample. Still, we find that about half of all households (i.e. 49%) are consistent with the GQL specification. For these households, we cannot reject the transferable utility hypothesis. By contrast, the QL utility specification appears to be overly stringent for the current data set: not a single household passes the corresponding rationalizability condition.

The remaining columns of Table 1 report on the power distribution for the two rationalizability tests. First, for the QL specification we obtain that the power distribution is almost entirely centered around unity, which reveals a (nearly) 100% probability of rejecting random behavior for each individual household. At this point, we note that this high power should not be too surprising given our previous finding that the QL condition is rejected for every house-

[11] We refer to Bronars (1987) and Andreoni and Harbaugh (2008) for a general discussion on alternative procedures to evaluate power in the context of revealed preference tests such as ours.

Figure 1: Power



hold in our sample. Next, we observe that the discriminatory power is rather substantially lower for the GQL test than for the QL test. However, the GQL test has reasonable power for a considerable part of the households; see, for example, the median, 3rd quartile and max values in Table 1. Figure 1 provides a corresponding kernel estimation of the GQL power distribution. A notable observation is that this distribution is bimodal with peaks around 0.15 and 0.5. Overall, our results reveal quite some variation in the power of the GQL test: it is fairly low for some household but considerably high for other households.[12]

Table 1: Pass rates and power

condition	pass rate	power					
		mean	min	1st quartile	median	3rd quartile	max
QL	0	1	0.998	1	1	1	1
GQL	0.488	0.398	0.101	0.241	0.449	0.515	0.674

As a further investigation, we consider two additional power distributions for the GQL test. Specifically, we compare the power distribution for the group of households that pass the test

[12]At this point, it is worth indicating that the power of revealed preference tests is driven essentially by price and income variation over consumption observations (see for example Bronars (1987)). Here, we note that we have a rotating panel of households, so that both price and income differences can explain the observed power variation across households.

(pass group) with the one for the group of households that fail the test (fail group). Table 2 gives the results. We observe a trade–off between power and pass rate for the GQL test: the power distribution for the fail group generally dominates the distribution for the pass group. Interestingly, however, this trade-off seems to be not very prevalent. For example, the minimal power for the pass group actually exceeds (albeit slightly) the one for the fail group, and the differences between the median, 3rd quartile and max values in Table 2 are rather small. In our opinion, this indicates that the better fit of the transferable utility model for the pass group can hardly be attributed to a lower power of the GQL test for this group (as compared to the fail group).

Table 2: Power for pass and fail groups

group	power					
	mean	min	1st quartile	median	3rd quartile	max
pass	0.352	0.105	0.174	0.388	0.496	0.659
fail	0.441	0.101	0.394	0.473	0.531	0.674

Predictive success. As an additional exercise, we compute a predictive success measure for the two conditions that we study. This measure was recently introduced and axiomatized by Beatty and Crawford (2011) and is based on an original proposal of Selten (1991). It combines the pass rate and power of a particular behavioral condition into a single metric: for each household, it subtracts 1 minus the power measure from the pass measure (1 or 0). As such, the predictive success measure can be interpreted as a power-adjusted pass rate. The measure is always situated between -1 and 1. Generally, a higher predictive success value then reveals a better empirical performance of the behavioral condition that is subject to testing. More specifically, a predictive success value that is close to -1 pertains to a household that fails the rationalizability condition (i.e. pass measure equals 0) even though the power of the test is low (i.e. close to 0). Conversely, a predictive success value close to 1 indicates a household that passes the condition (i.e. pass measure equals 1) in a situation where this condition has high power (i.e. close to 1). Finally, a predictive success value that equals exactly zero means that the condition is not informative for the household at hand: the condition does not perform any better than the (uninformative) assumption that households exhibit random consumption behavior (for which the power is 0 and the pass measure equals 1, by construction). For a given household, we will use this zero value as a natural threshold value to identify a rationalizability condition as a 'bad' condition (if predictive success is below zero) or a 'good' condition (if predictive success is above zero).

Table 3 presents summary statistics for the predictive success measures that are relevant here. These statistics tell us about the empirical performance of the two rationalizability conditions at the aggregate level of our sample (with 1585 households). As a first observation, we note that the distribution is centered around zero for the QL condition, with (almost) no variation across observations. In fact, we could have expected this result on the basis of the 0% pass rate and (nearly) 100% power results that we presented before. Given the above, this suggests that the QL condition is not informative for the data at hand. By contrast, the pattern seems to be more indicative for the GQL condition. Like for the QL condition, we again get that the mean predictive success score is close to zero, but now there is more variation across households.

Table 3: Predictive success

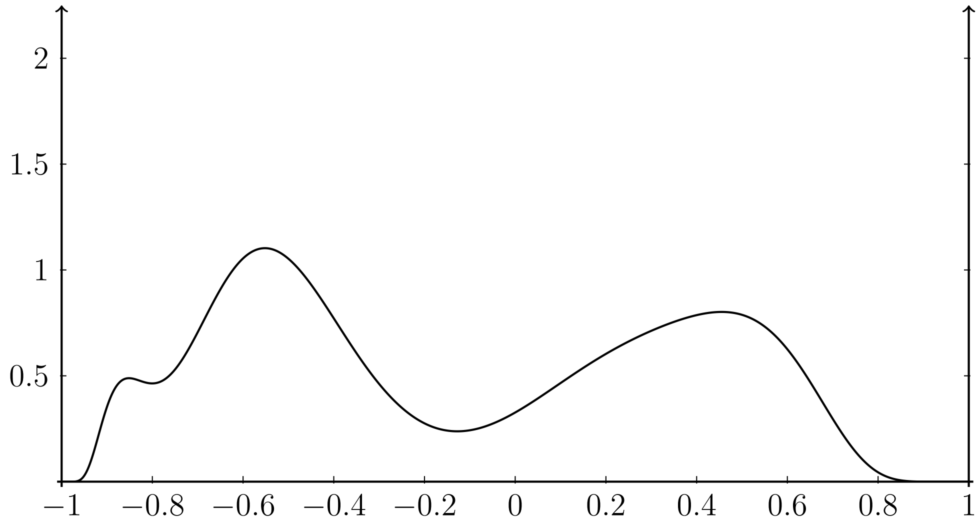
condition	mean	min	1st quartile	median	3rd quartile	max
QL	0	-0.002	0	0	0	0
GQL	-0.114	-0.899	-0.530	-0.371	0.376	0.659

To provide a better view of this cross-household variation, Figure 2 depicts an estimation of the predictive success distribution for the GQL condition. Interestingly, the figure reveals a clear bimodal pattern: the distribution achieves a first peak around -0.55 and a second peak around 0.5. One possible interpretation of this bimodalility is that it suggests a particular split-up of our original sample of households: for a substantial group of households the GQL condition can be identified empirically as a 'good' one, whereas it is a rather 'bad' condition for the remaining households. In turn, this indicates that the adequacy of the GQL condition may depend on the specific household (characteristics) at hand. In this respect, we have compared observable characteristics (in our data set) for the two household groups, i.e. households with predictive success above zero (for which GQL is a 'good' condition) and households with predictive success below zero (for which GQL is a 'bad' condition). We considered the following characteristics: age of the household head, number of household members, specialized worker occupation and home ownership. However, we found no statistical differences between the two subsamples. A (tentative) conclusion can be that other (unobserved) household characteristics drive the adequacy of the GQL condition.

Measurement error. So far, we have considered basic tests for the GQL and QL conditions. As we mentioned before, these are sharp tests that do not account for measurement error. They only tell us whether households are exact optimizers in terms of a particular behavioral (GQL or QL) condition for the data at hand. Obviously, this is a demanding premise since consumption data are often contaminated by (small) error. From this perspective, it seems useful to complement our first assessment with a second exercise which, for household behavior violating the sharp rationalizability conditions, evaluates how close observed behavior is to rationalizability once we account for measurement error. Therefore, in what follows we concentrate on an extension of the basic tests that seeks minimal data adjustments needed for exact consistency with the behavioral conditions under study. For our specific application, we here focus on adjustments of the price data. This follows Crawford (2010), who motivates the possibility of price errors for the sample at hand.[13] Our following method for dealing with (price) errors

[13]However, it should be clear that our following method is easily accommodated to account for quantity errors in (other) applications where this seems more adequate.

Figure 2: Predictive success



essentially adapts an idea of Varian (1985), who originally focused on revealed preference analysis of the unitary model. After introducing the method, we will present (and compare) our empirical results for the GQL and QL specifications.

For each observation t, we let the vectors $\boldsymbol{\varepsilon_t}$ and $\boldsymbol{\upsilon_t}$ represent the measurement errors in the prices of, respectively, the private goods ($\mathbf{p}_t$) and the public goods ($\mathbf{P}_t$). We assume that these errors are independently normally distributed with zero mean and variance σ^2; Γ_t and Λ_t stand for the diagonal matrices with as diagonal entries $1 + \varepsilon_{t,n}$ and $1 + \upsilon_{t,k}$.[14] Then, we have that the observed prices $\mathbf{p}_t$ and $\mathbf{P}_t$ are related to the 'true' (but unobserved) prices $\tilde{\mathbf{p}}_t$ and $\tilde{\mathbf{P}}_t$ in the following way:

$$\tilde{\mathbf{p}}_t = \Gamma_t \mathbf{p}_t \qquad \text{and} \qquad \tilde{\mathbf{P}}_t = \Lambda_t \mathbf{P}_t, \tag{2}$$

We note that we assume a multiplicative error structure. This allows for a higher variance associated with higher prices, which is intuitively plausible. It will also simplify the interpretation of our following empirical results.

Using (2), we can reformulate our revealed preference conditions in terms of the true prices $\tilde{\mathbf{p}}_t$ and $\tilde{\mathbf{P}}_t$, by replacing the conditions (IP.1) and (IP.2) with, respectively,

$$U_t^B - U_v^B \leq \Gamma_v \mathbf{p}_v (\mathbf{q}_t - \mathbf{q}_v) + \mathbf{P}_v^B (\mathbf{Q}_t - \mathbf{Q}_v), \tag{IP.1-e}$$

$$\mathbf{P}_v^A + \mathbf{P}_v^B = \Lambda_v \mathbf{P}_v. \tag{IP.2-e}$$

Now, if consumption is not consistent with the conditins (IP.1)–(IP.5), we know that we will need $\boldsymbol{\varepsilon_t}$ and $\boldsymbol{\upsilon_t}$ different from zero to satisfy the conditions (IP.1-e), (IP.2-e) and (IP.3)-(IP.5). As we do not now the true values of the errors $\boldsymbol{\varepsilon_t}$ and $\boldsymbol{\upsilon_t}$ (because the true prices are

[14]Here, $\varepsilon_{t,n}$ stands for the n-th entry of the vector $\boldsymbol{\varepsilon_t}$ and $\upsilon_{t,k}$ for k-th entry of the vector $\boldsymbol{\upsilon_t}$.

unobserved), we compute the minimal sum of squared errors that makes the data satisfy these last conditions. This can effectively be interpreted in terms of the minimal price adjustments needed for obtaining consistency with the behavioral condition subject to evaluation; it quantifies how close observed behavior is to such consistency. Formally, we define $\widehat{\boldsymbol{\varepsilon}}_t$ and $\widehat{\boldsymbol{v}}_t$ that minimize

$$\widehat{\mathfrak{R}} = \sum_{t \in T} \left(\widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t + \widehat{\boldsymbol{v}}_t \widehat{\boldsymbol{v}}_t \right)$$

subject to the constraints (IP.1-e), (IP.2-e) and (IP.3)–(IP.5). This obtains a minimization problem with a quadratic objective function and mixed linear programming constraints. Again, solution methods for such problems are fairly standard.

Interestingly, the outcome of the minimization problem can also be used to statistically test the null hypothesis that rationalizability holds when accounting for measurement error. To see this, let $\mathfrak{R}$ represent the sum of squared errors associated with the true prices (i.e. $\mathfrak{R} = \sum_{t \in T} \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t + \boldsymbol{v}_t \boldsymbol{v}_t \right)$). Given our above assumptions, we know that[15]

$$\frac{\mathfrak{R}}{\sigma^2} \sim \chi^2_{120}.$$

Thus, for any critical value C_α taken from the given chi-squared distribution, we reject the null hypothesis of rationalizability at a significance level α as soon as $\mathfrak{R}/\sigma^2 \geq C_\alpha$.

Of course, in practice we do not observe $\mathfrak{R}$ or σ^2. However, we can use that, by construction, $\widehat{\mathfrak{R}}$ will bound $\mathfrak{R}$ from below, so that (for given σ^2)

$$\frac{\widehat{\mathfrak{R}}}{\sigma^2} \leq \frac{\mathfrak{R}}{\sigma^2}.$$

Then, following an original idea of Varian (1985), we can compute (for critical value C_α associated with some predefined significance level α)

$$\widehat{\sigma} = \sqrt{\frac{\widehat{\mathfrak{R}}}{C_\alpha}}.$$

Our above argument directly obtains that $\widehat{\sigma}$ gives the standard deviation (of the measurement errors $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{v}_t$) one minimally needs to account for in order not to reject the null hypothesis that behavior is rationalizable. In practice, $\widehat{\sigma}$ can be compared to one's prior belief regarding the true standard deviation σ. If $\widehat{\sigma}$ exceeds this prior belief, then one can reject the null hypothesis for rationalizability (at the significance level α), and vice versa. It is worth pointing out that, clearly, this test procedure is fairly conservative; the null hypothesis of rationalizable behavior will be rejected only if there is strong evidence against it.

Table 4 provides some key values for the empirical distribution of the measure $\widehat{\sigma}$ associated with significance level $\alpha = 0.05$. It reports for the subset of households that fail the 'sharp' GQL and QL tests. Recall that we used a multiplicative error structure, which, from an interpretation point of view, implies that $\widehat{\sigma}$ captures standard deviation of the price errors in percentage terms.

[15]The degrees of freedom equal the number of data points that are perturbed. In our application, we consider 8 observations and 15 goods, which makes that we have 120 prices for which we allow errors.

Table 4: Measurement error ($\hat{\sigma}$)

	mean	min	1st quartile	median	3rd quartile	max
QL	0.0057	0.0005	0.0036	0.0055	0.0074	0.0158
GQL	0.0017	0	0.0002	0.0011	0.0026	0.0111

We first consider the GQL results. Here, we conclude that measurement errors may effectively explain the rejections of the 'sharp' test. Indeed, the maximum $\widehat{\sigma}$ value amounts to no more than $1,1\%$. In other words, the behavior of this 'worst fitting' household can be rationalized in terms of transferable utility (or GQL utility) as soon as we believe the price errors have a standard deviation of (minimally) $1,1\%$. In addition, looking at the 3rd quartile value, we conclude rationalizability of 75% of the households if we accept a standard deviation of only 0.3%. These are obviously very small numbers. As such, this provides a strong empirical case for transferable utility in the case one believes prices are measured with errors.

Let us then focus on the QL results. As a first observation, we find that the corresponding values in Table 4 are everywhere above those for the GQL model. This falls in line with the basic test results that we reported before, and so again suggests that the GQL specification has better empirical support than he QL specification. However, the $\widehat{\sigma}$ values are generally low. For example, the maximum value in the sample amounts to only 1.5%. Thus, if one believes prices are measured with error, the observed consumption behavior can be rationalized fairly well in terms of the QL specification; a little degree of price adjustments allows for such a rationalization. From this perspective, we may conclude that there does not seem to be a very strong empirical case against the QL specification when accounting for price errors.

What do we learn from all this? Our application allows for drawing both methodological and empirical conclusions. At the methodological level, we believe that our application convincingly demonstrates the practical usefulness of our revealed preference characterization for assessing the validity of transferable (or GQL) utility in real life settings. Also, it shows that the conditions we derived above provide a useful basis for analyzing power and predictive success, and are directly extended to account for measurement error in the consumption data. More generally, this illustrates that using our integer programming formulation does not restrict the revealed preference analysis. Starting from this formulation, we can empirically address the same type of questions (including methodological extensions) as in more standard analysis of the unitary model of household consumption. In this respect, we also refer to our discussion in the concluding section, where we touch upon recovery and forecasting analysis.

Next, at the empirical level, we have investigated the validity of the transferable utility hypothesis, and hereby compared the performance of the GQL and QL specifications for a particular sample of Spanish households. We conducted two complementary exercises. The conclusions of our first exercise, which focused on the basic tests without measurement error, are quite clear-cut. A first main conclusion is that the GQL specification is more useful than the QL specification, which is strongly rejected for our data. In fact, for a considerable subset of households the GQL specification performs rather well empirically. However, if we focus on the 'sharp' GQL test, there is also a substantial amount of households that behave inconsistently with transferable utility. We were not able to explain the observed violations of transferable utility by characteristics observed in our data set. So, from this point of view, other (unobserved) characteristics may define (exact) consistency with the transferable utility specification.

Our second exercise took a very different perspective and investigated whether measurement error in the price data can explain the observed violations of (GQL and QL) rationalizability. Interestingly, we did find that accounting for a little measurement error can rationalize all household behavior in terms of transferable utility (i.e. obtains consistency with the GQL condition). In fact, and somewhat remarkably, we now also obtained a much stronger empirical case pro the QL specification: when considering a small amount of measurement error, we can make the observed consumption behavior consistent with the QL specification under small adjustments of the observed prices. Thus, our earlier conclusion in favor of the GQL specification and against the QL specification seems to depend largely on the maintained assumption that prices are measured accurately.

6 Conclusion

We have presented revealed preference conditions that must be satisfied by observed behavior to be consistent with transferable utility (or GQL utility) under Pareto efficiency. These conditions are easily verified by using integer programming techniques, which is attractive from a practical point of view. This provides an easy-to-apply framework for evaluating the empirical realism of the transferable utility hypothesis in observational settings. As a side-result, our theoretical discussion also made clear how the transferable utility model is situated 'between' the quasi-linear (QL) and unitary model: its (revealed preference) testable implications are weaker than the QL implications but stronger than the unitary implications.

We have demonstrated the usefulness of our revealed preference framework by an empirical application to Spanish households. Generally, our results suggest that the assumption of transferable utility is a useful one for this sample of households. First, when considering our basic revealed preference tests, which abstract from errors in the consumption data, we concluded that a large class of households satisfies transferable utility. We then investigated whether accounting for errors in the price data allowed for rationalizing behavior that is not exactly consistent with transferable utility. Interestingly, we found that this is effectively the case: we needed only small adjustments of the observed prices to rationalize all household consumption behavior in terms of transferable utility. It seems interesting to investigate whether these conclusions are confirmed for other household data (that potentially imply more powerful revealed preference tests). We provided a framework that allows for doing so in a fairly easy manner.

In our application, we also compared the empirical fit of the GQL specification with the one of the more restrictive QL specification. For the basic revealed preference tests, we found quite an important difference: none of the households appeared to be (exactly) consistent with QL utility, while a substantial part of the households meet the conditions for GQL utility. This difference became much less pronounced when we accounted for price errors. We found that small price adjustments can obtain rationalizability of the observed household behavior in terms of the QL specification. This suggests that, for the sample under study, the favorable results for the GQL specification as compared to the QL specification are mainly driven by the assumption that prices are measured without error. At this point, however, it is also worth to recall that the test procedure we used to account for measurement error is a fairly conservative one.

We see different avenues for follow-up research. From an empirical point of view, an in-depth investigation can provide more detailed insights into possible explanations of violations of transferable utility. For example, one may use richer data sets than the one we studied (e.g. involving additional consumption observations and/or household characteristics) to more thoroughly investigate the specific household characteristics that define the adequacy of the transferable utility model. We believe this is particularly interesting given the wide use of the transferable utility hypothesis in (theoretical) household economics. Next, as also indicated in the Introduction, our framework can equally be used for assessing the validity of the transferable utility hypothesis in alternative (non-household) settings where this assumption crucially underlies important theoretical results.

Finally, to keep our exposition simple, our analysis has concentrated on the characterization of transferable utility, and testing consistency of observed behavior with this characterization. If observed behavior is found consistent with a behavioral hypothesis, then natural next questions involve recovering/identifying the corresponding decision model that rationalizes the observed consumption behavior, and to forecasting behavior in new situations. In this respect, it is worth emphasizing that our revealed preference characterization does allow for subsequent recovery and forecasting analysis. For example, this analysis can develop along the lines of Varian (1982) and, more recently, Blundell, Browning, and Crawford (2008), who use similar revealed preference methods to consider such questions for the unitary model. In this respect, we recall from our theoretical discussion that the GQL (or transferable utility) model has stronger testable implications than the unitary model, which is usually considered in revealed preference applications. As such, we can expect that using the GQL specification (when it cannot be rejected) can effectively produce more vigorous recovery and forecasting results.

References

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

Andreoni, J., Harbaugh, W., 2008. Power indices for revealed preference tests. Tech. Rep. 2005-10, University of Winconsin-Madison Department of Economics.

Beatty, T. K. M., Crawford, I. A., 2011. How demanding is the revealed preference approach to demand. American Economic Review 101, 2782–2795.

Becker, G. S., 1974. A theory of social interactions. Journal of Political Economy 82, 1063–1093.

Bergstrom, T., 1989. A fresh look at the rotten-kid theorem and other household mysteries. Journal of Political Economy 97, 1138–1159.

Bergstrom, T., 1997. The Handbook of Population and Family Economics. Ch. A Survey of Theories of the Family, pp. 22–74.

Bergstrom, T. C., Cornes, R. C., 1981. Gorman and Musgrave are dual: an antipodean theorem on public goods. Economics Letters 7, 371–378.

Bergstrom, T. C., Cornes, R. C., 1983. Independence of allocative efficiency from distribution in the theory of public goods. Econometrica 51, 1753–1765.

Blundell, R., Browning, M., Crawford, I., 2008. Best nonparametric bounds on demand responses. Econometrica 76, 1227–1262.

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

Brown, D. J., Calsamiglia, C., 2007. The nonparametric approach to applied welfare analysis. Economic Theory 31, 183–188.

Browning, M., Chiappori, P.-A., Weiss, Y., 2011. Family Economics. Cambridge University Press, forthcoming.

Browning, M., Collado, M. D., 2001. The response of expenditures to anticipated income changes: Panel data estimates. American Economic Review 91, 681–692.

Cherchye, L., Crawford, I., De Rock, B., Vermeulen, F., 2009a. Quantifying Consumer Preferences: Estimating Demand Systems. Emerald Press, Ch. The Revealed Preference approach to Demand.

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

Cherchye, L., De Rock, B., Vermeulen, F., 2009b. Opening the black box of intra-household decision-making. Journal of Political Economy 117, 1074–1104.

Cherchye, L., De Rock, B., Vermeulen, F., 2011. The revealed preference approach to collective consumption behavior: Testing and sharing rule recovery. Review of Economic Studies 78, 176–198.

Cherchye, L., Demuynck, T., De Rock, B., 2010. Noncooperative household consumption with caring. CES DP 10.34, K.U.Leuven.

Chiappori, P., 1988. Rational household labor supply. Econometrica 56, 63–89.

Chiappori, P.-A., 2010. Testable implications of transferable utility. Journal of Economic Theory 145, 1302–1317.

Coase, R. H., 1960. The problem of social cost. Journal of Law and Economics 3, 1–44.

Crawford, I., 2010. Habits revealed. Review of Economic Studies 77, 1382–1402.

Davis, M., Maschler, M., 1965. The kernel of a cooperative game. Naval Research Logistics Quarterly 12, 223–259.

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

Houthakker, H. S., 1950. Revealed preference and the utility function. Economica 17, 159–174.

Hurwicz, L., 1995. What is the coase theorem? Japan and the World Economy 7, 49–74.

Samuelson, P. A., 1938. A note on the pure theory of consumer's behavior. Economica 5, 61–71.

Schmeidler, D., 1969. The nucleolus of a characteristic function game. SIAM Journal on Applied Mathematics 17, 1163–1170.

Selten, R., 1991. Properties of a measure of predictive success. Mathematical Social Sciences 21, 153–167.

Shapley, L., Shubik, M., 1972. The assignment game I: The core. International Journal of Game Theory 1, 111–130.

Shapley, L. S., 1953. Contributions to the Theory of Games II. Princeton University Press, Ch. A Value for n-Person Games, pp. 307–317.

Varian, H., 1982. The nonparametric approach to demand analysis. Econometrica 50, 945–974.

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

Varian, H. R., 1988. Revealed preference with a subset of goods. Journal of Economic Theory 46, 179–185.

Appendix A: proofs

Proof of Proposition 1

(2 $\rightarrow$ 3). By convexity of the function $\alpha(\mathbf{Q})$ and concavity of the function $\beta(\mathbf{q}, \mathbf{Q})$ we must have that for all observations $t, v \in T$:

$$\alpha(\mathbf{Q}_t) - \alpha(\mathbf{Q}_v) \geq \frac{\partial \alpha(\mathbf{Q}_v)}{\partial \mathbf{Q}} (\mathbf{Q}_t - \mathbf{Q}_v),$$

$$\beta^m(\mathbf{q}_t, \mathbf{Q}_t) - \beta^m(\mathbf{q}_v, \mathbf{Q}_v) \leq \frac{\partial \beta(\mathbf{q}_v, \mathbf{Q}_v)}{\partial \mathbf{q}} (\mathbf{q}_t - \mathbf{q}_v) + \frac{\partial \beta(\mathbf{q}_v, \mathbf{Q}_v)}{\partial \mathbf{Q}} (\mathbf{Q}_t - \mathbf{Q}_v).$$

For all $t \in T$, define $\alpha_t = \alpha(\mathbf{Q}_t)$, $\beta_t = \beta(\mathbf{q}_t, \mathbf{Q}_t)$, $\bar{u}_t = u(x_t, \mathbf{q}_t, \mathbf{Q}_t)$, $\boldsymbol{\lambda}_t^{\alpha} = \dfrac{\partial\alpha(\mathbf{Q}_t)}{\partial\mathbf{Q}}$ and $\boldsymbol{\lambda}_t^{\beta} = \dfrac{\partial\beta(\mathbf{q}_t, \mathbf{Q}_t)}{\partial\mathbf{Q}}$. Then, substituting and using the first order conditions (foc.1)-(foc.3) obtains conditions (RP.1)-(RP.4).

($1 \to 3$) The proof is similar to the case ($2 \to 3$) except now, we define $\beta_t = \sum_m \beta^m(\mathbf{q}_t^m, \mathbf{Q}_t)$ and $\boldsymbol{\lambda}_t^{\beta} = \sum \dfrac{\partial\beta(\mathbf{q}_t^m, \mathbf{Q}_t)}{\partial\mathbf{Q}}$.

($3 \to 2$). Define the functions $\alpha(\mathbf{Q})$ and $\beta(\mathbf{q}, \mathbf{Q})$ in the following way:

$$\alpha(\mathbf{Q}) = \max_{t \in T} \left\{\alpha_t + \boldsymbol{\lambda}_t^{\alpha} (\mathbf{Q} - \mathbf{Q}_t)\right\}, \tag{A.1}$$

$$\beta(\mathbf{q}, \mathbf{Q}) = \min_{t \in T} \left\{\beta_t + \mathbf{p}_t(\mathbf{q} - \mathbf{q}_t) + \boldsymbol{\lambda}_t^{\beta} (\mathbf{Q} - \mathbf{Q}_t)\right\}. \tag{A.2}$$

Define $u(x, \mathbf{q}, \mathbf{Q}) = \dfrac{x}{\alpha(\mathbf{Q})} + \dfrac{\beta(\mathbf{q}, \mathbf{Q})}{\alpha(\mathbf{Q})}$.

The function α is convex and β is concave, hence u is quasi-concave. Further, it is increasing in both $\mathbf{q}$ and $\mathbf{Q}$. Finally, using a similar argument as Varian (1982, p.970), we can derive that $\alpha(\mathbf{Q}_t) = \alpha_t$ and $\beta(\mathbf{q}_t, \mathbf{Q}_t) = \beta_t$ for all $t \in T$

Given all this, we can prove the result ad absurdum. Suppose that S is not TU-rationalizable. Then, there must exist an allocation $\{x, \mathbf{q}, \mathbf{Q}\}$ such that $x + \mathbf{p}_t\mathbf{q} + \mathbf{P}_t\mathbf{Q} < x_t + \mathbf{p}_t\mathbf{q}_t + \mathbf{P}_t\mathbf{Q}_t$ and $u(x, \mathbf{q}, \mathbf{Q}) \geq u(x_t, \mathbf{q}_t, \mathbf{Q}_t) = \bar{u}_t$. We thus get

$$\begin{aligned} x + \mathbf{p}_t\mathbf{q} + \mathbf{P}_t\mathbf{Q} &\geq \bar{u}_t\alpha(\mathbf{Q}) - \beta(\mathbf{q}, \mathbf{Q}) + \mathbf{p}_t\mathbf{q} + \mathbf{P}_t\mathbf{Q} \\ &\geq \bar{u}_t\alpha_t - \beta_t + \left(\boldsymbol{\lambda}_t^{\alpha}\bar{u}_t - \boldsymbol{\lambda}_t^{\beta}\right)(\mathbf{Q} - \mathbf{Q}_t) - \mathbf{p}_t(\mathbf{q} - \mathbf{q}_t) + \mathbf{p}_t\mathbf{q} + \mathbf{P}_t\mathbf{Q} \\ &= x_t + \mathbf{p}_t\mathbf{q}_t + \mathbf{P}_t\mathbf{Q}_t, \end{aligned}$$

which gives the wanted contradiction. (The first inequality combines $u(x, \mathbf{q}, \mathbf{Q}) = (x/\alpha(\mathbf{Q})) + (\beta(\mathbf{q}, \mathbf{Q})/\alpha(\mathbf{Q}))$ with $u(x, \mathbf{q}, \mathbf{Q}) \geq \bar{u}_t$, the second inequality uses (A.1) and (A.2), and the final equality uses (RP.3) and (RP.4).)

($3 \to 1$) The argument is similar to one for ($3 \to 2$), when using the additional definition $\beta^m(\mathbf{q}^m, \mathbf{Q}) = \dfrac{1}{M}\beta(M\mathbf{q}^m, \mathbf{Q})$. Then, for all $t \in T$ and $m \leq M$, we set $\mathbf{q}_t^m = \mathbf{q}_t/M$ and $x_t^m = x_t/M$.

Proof of Proposition 2

($1 \to 2$) Assume that there exist numbers x_t such that $\{\mathbf{p}_t, \mathbf{P}_t; x_t, \mathbf{q}_t, \mathbf{Q}_t\}_{t \in T}$ is TU-rationalizable. Then, it follows from Proposition 1 that there exist positive numbers α_t, β_t and $\bar{u}_t$, vectors

$\boldsymbol{\lambda}_t^\alpha \in \mathbb{R}_-^K$ and $\boldsymbol{\lambda}_t^\beta \in \mathbb{R}_{++}^K$ such that

$$\alpha_t - \alpha_v \geq \boldsymbol{\lambda}_v^\alpha(\mathbf{Q}_t - \mathbf{Q}_v) \tag{RP.1}$$
$$\beta_t - \beta_v \leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + \boldsymbol{\lambda}_v^\beta(\mathbf{Q}_t - \mathbf{Q}_v) \tag{RP.2}$$
$$\boldsymbol{\lambda}_t^\beta - \boldsymbol{\lambda}_t^\alpha \bar{u}_t = \mathbf{P}_t \tag{RP.3}$$
$$\bar{u}_t = \frac{x_t}{\alpha_t} + \frac{\beta_t}{\alpha_t} \tag{RP.4}$$

Setting, for all $t \in T$, $\beta_t = U_t^B$, $\boldsymbol{\lambda}_t^\beta = \mathbf{P}_t^B$ and $\mathbf{P}_t^A = -\boldsymbol{\lambda}_t^A \bar{u}_t$ translates condition (RP.2) and (RP.3) into conditions (RP.6) and (RP.7). So we only need to demonstrate condition (RP.5).

Multiplying (RP.1) by minus one, gives:

$$-\alpha_t - (-\alpha_v) \leq \frac{1}{\bar{u}_t} \mathbf{P}_v^A (\mathbf{Q}_t - \mathbf{Q}_v)$$

Given this, setting $\lambda_t^A = 1/\bar{u}_t > 0$ and $U_t^A = -\alpha_t - \min_v\{-\alpha_v\} \geq 0$ establishes condition (RP.5).

($2 \rightarrow 1$) Assume that there exist numbers U_t^A, U_t^B and λ_t^A, and vectors $\mathbf{P}_v^A$ and $\mathbf{P}_v^B$ such that

$$U_t^A - U_v^A \leq \lambda_t^A \left[\mathbf{P}_v^A(\mathbf{Q}_t - \mathbf{Q}_v)\right] \tag{RP.5}$$
$$U_t^B - U_v^B \leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + \mathbf{P}_v^B(\mathbf{Q}_t - \mathbf{Q}_v) \tag{RP.6}$$
$$\mathbf{P}_t^A + \mathbf{P}_t^B = \mathbf{P}_t \tag{RP.7}$$

First, by setting, for all $t \in T$, $\beta_t = U_t^B$, $\boldsymbol{\lambda}_t^\beta = \mathbf{P}_t^B$, we derive (RP.2).

Next, we define $\bar{u}_t = 1/\lambda_t^A$ and $\mathbf{P}_t^A/\bar{u}_t = -\boldsymbol{\lambda}_t^\alpha$. Substitution in condition (RP.7) gives condition (RP.3).

Further, multiplying (RP.5) by minus one gives,

$$-U_t^A - (-U_v^A) \geq \boldsymbol{\lambda}_t^\alpha (\mathbf{Q}_t - \mathbf{Q}_v) \tag{A.3}$$

As $\bar{u}_t > 0$, there exist a number $\delta > 0$ such that $\bar{u}_t > \delta$ for all $t \in T$. Now, consider a number $z \in \mathbb{R}_{++}$ and define α_t such that (i) $\alpha_t \equiv -U_t^A + z > 0$ ($\forall t \in T$) and (ii) $0 < \beta_t/\alpha_t \leq \delta$. These conditions can be guaranteed by taking z large enough. Using this definition of α_t in condition (A.3) above gives condition (RP.1).

Finally, we define x_t such that

$$x_t \equiv \alpha_t \bar{u}_t - \beta_t > 0,$$

which obtains condition (RP.4).

Proof of Proposition 3

This result uses the equivalence of (RP.5) and GARP; this is stated more formally in Theorem 1 below. Next, in the main text we argued that (IP.3)-(IP.5) do allow for verifying GARP for our setting.

Appendix B: Conditions (RP.1)-(RP.4) imply GARP

From (RP.2) it follows that

$$(\beta_t + x_t) - (\beta_v + x_v) \leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + \boldsymbol{\lambda}_v^\beta(\mathbf{Q}_t - \mathbf{Q}_v) + x_t - x_v.$$

Then, using (RP.4) we obtain

$$\bar{u}_t\alpha_t - \bar{u}_v\alpha_v \leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + (\lambda)_v(\mathbf{Q}_t - \mathbf{Q}_v) + x_t - x_v.$$

Next, adding to both sides $\bar{u}_v(\alpha_v - \alpha_t)$ and making use of (RP.1) gives

$$\begin{aligned}(\bar{u}_t - \bar{u}_v)\alpha_t &\leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + \boldsymbol{\lambda}_v^\beta(\mathbf{Q}_t - \mathbf{Q}_v) + \bar{u}_v(\alpha_v - \alpha_t) + x_t - x_v \\ &\leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + \boldsymbol{\lambda}_v^\beta(\mathbf{Q}_t - \mathbf{Q}_v) - \bar{u}_v\boldsymbol{\lambda}_v^\alpha(\mathbf{Q}_t - \mathbf{Q}_v) + x_t - x_v.\end{aligned}$$

Finally, from (RP.3) we get

$$(\bar{u}_t - \bar{u}_v)\alpha_t \leq \mathbf{p}_v(\mathbf{q}_t - \mathbf{q}_v) + \mathbf{P}_v(\mathbf{Q}_t - \mathbf{Q}_v) + x_t - x_v. \tag{3}$$

Now, the above inequality shows that, if $\mathbf{p}_v\mathbf{q}_v + \mathbf{P}_v\mathbf{Q}_v + x_v \geq \mathbf{p}_v\mathbf{q}_t + \mathbf{P}_v\mathbf{Q}_t + x_t$, then $\bar{u}_v \geq \bar{u}_t$. Hence, if $(\mathbf{q}_v, \mathbf{Q}_v, x_v)R(\mathbf{q}_t, \mathbf{Q}_t, x_t)$, then also $\bar{u}_v \geq \bar{u}_t$. As such, if on the contrary GARP is not satisfied, there must exist observations t and $v \in T$ such that $\bar{u}_v \geq \bar{u}_t$ and $\bar{u}_t > \bar{u}_v$, a contradiction.

Appendix C: Afriat's Theorem

In the main text, we make use of Afriat's Theorem. This result was stated by Varian (1982) and is based on the original work of Afriat (1967). It is probably the single most important theorem in the revealed preference literature. To facilitate our exposition in the main text, we briefly recapture the result here. We refer to Varian (1982) for a more detailed discussion.

Let us consider a general setting with a price-quantity set Z as introduced in Section 5 of the main text. We consider the following rationalizability concept:

Definition 3 (U-rationalizable) *The set $Z = \{\mathbf{w}_l; \mathbf{y}_l\}_{l\in L}$ is utility (U)-rationalizable if there exist a non-satiated utility function u such that each quantity bundle $\mathbf{y}_l$ maximizes the function u in the following sense: $\mathbf{y}_l \in \arg\max_{\mathbf{y}} u(\mathbf{y})\ s.t.\ \mathbf{w}_l\mathbf{y} \leq \mathbf{w}_l\mathbf{y}_l$.*

We can now state Afriat's Theorem.

Theorem 1 (Afriat's Theorem) *Consider a set $Z = \{\mathbf{w}_l; \mathbf{y}_l\}_{l\in L}$. The following conditions are equivalent:*

1. *The set Z is U-rationalizable;*

2. *The set Z satisfies GARP;*

3. *For all $l \in L$, there exist $U_l \in \mathbb{R}_+$ and $\lambda_l \in \mathbb{R}_{++}$ such that, for all $l, k \in L$:*

$$U_l - U_k \leq \lambda_k \mathbf{w}_k(\mathbf{y}_l - \mathbf{y}_k).$$

4. *There exist a strictly increasing, continuous and concave utility function that provides a U-rationalization for Z.*

In Section 5 of our main text, we use two important implications of this result. First, the equivalence between statements 1 (or 4) and 2 implies that a price-quantity set Z is U-rationalizable by some utility function if and only if it is consistent with GARP. Second, the equivalence between statements 2 and 3 means that the set Z is consistent with GARP if and only if it satisfies a number of inequalities defined in the unknowns U_l and λ_l. These last inequalities are commonly referred to as 'Afriat inequalities' corresponding to the set Z. Intuitively, these Afriat inequalities allow us to obtain estimates for the utility levels (U_l) and marginal utilities (λ_l) attained at each l whenever the set Z is rationalizable.