# An Application of Behavioral Welfare Economics to Consumption Data

Preliminary and incomplete

Elias Bouacida*
Université Paris 1 Panthéon Sorbonne – Paris School of Economics
Daniel Martin
Kellog School of Management, Northwestern University

August 28, 2015

**Abstract**

Revealed preferences are a potentially powerful tool for conducting welfare analyses. However, due to forces such as behavioral biases, "irrational behaviors" such as non-transitivity (i.e. cycles of preferences) can appear in – revealed – preferences and hampers their role for welfare analysis. In this work, we provide the first empirical application of a leading proposal for welfare analysis in the presence of behavioral biases, introduced in Bernheim and Rangel (2009) and called the Strict Unambiguous Choice Relation (hereafter Strict ucr). Two potential problems emerge in the application of this approach: Strict ucr can be cyclic on standard consumption data. Strict ucr is very conservative and may be overly incomplete and thus, uninformative (Rubinstein and Salant 2012). We address these concerns using a dataset of package grocery purchases by 397 households over a two years period. We find that Strict ucr is cyclic on incomplete data, but much less than revealed preferences (only 8% of households exhibit cycles, to as much as 99% for revealed preferences). Also, we find that Strict ucr is not overly incomplete, as in our application almost all bundles are comparable. We introduce a new relation strengthening Strict ucr guaranteeing acyclicity. It is shown to be equivalent to Strict ucr on complete data. We finally show that Strict ucr makes tight out of sample predictions.

Keywords: Behavioral economics, Economic welfare, Weak Axiom of Revealed Preferences, Scanner data
Jel Classification: D60, D110

*Corresponding author: elias.bouacida@psemail.eu

# Contents

# 1 Introduction

The main aim of this paper is to provide the first of application of a theoretical preference relation introduced by Bernheim and Rangel (2009), that lead to a new trend in the literature, "behavioral welfare economics". These terms summarize their purpose: providing theoretical foundations to welfare inferences in environments with behavioral biases. In order to do so, they use the old apparatus of revealed preferences and mix it with new insights from behavioral economics. They are in particular concerned by the apparent lack of theoretical foundations of welfare inferences in behavioral economics: for each bias, a model is introduced to explain it. The construction of those models is often criticized as being very *ad hoc*: the modeler has a lot of freedom in choosing what to take into account and what is right or wrong, which is not very satisfactory.

## 1.1 Behavioral welfare literature

Behavioral economics grew after the seminal paper of Tversky and Kahneman (1974) in response the empirical failure of the classical model of utility maximization and in particular to the failure of the revealed preferences approach first sketched by Samuelson in "A Note on the Pure Theory of Consumer's behaviour"[1]. The empirical failure of the axiom of revealed preferences. Afriat (1967) demonstrated how to apply the Strong Axiom of Revealed Preferences (hereafter sarp) to finite datasets and provided an algorithm to build a utility out of the observations. It should be noted that the test of various axioms of revealed preferences is a closed question: the answer is either yes, the data satisfies the axiom, or no, it does not. As such, economists have tried to assess how far the observation are from verifying the axiom, a prominent example is the *Afriat Efficiency Index*.

Failures of revealed preferences led economists to try to assess preferences from other standpoints. These trials use behavioral economics to provide new insights for welfare inferences. At first, for each bias economists built a new model of decision and thus new welfare inferences (left-digit bias, *satisficing*, reference dependence and so on). However, this approach is not very robust nor general. It does not allow to work with several biases at the same time, as it is likely to be the case when looking at consumption data, the kind of data we are interested in here. Another drawback is the risk of arbitrariness from the modeler: each specific model can be tailored more or less at will, which might lead to different conclusions.

To alleviate those problems, during the last decade, economists have tried to build more consistent and general models to infer welfare in the presence of biases. The first (general) attempts to provide welfare inferences from choice data, using behavioral economics were proposed by Salant and Rubinstein (2008) and Bernheim and Rangel (2009). These authors separately reached the same idea of using frames, as defined by Tversky and Kahneman (1981). The main idea is to incorporate an ancillary condition (=a frame), which is not relevant for welfare inference but is relevant to understand the ob-

[1] We will not developed on the history of revealed preferences here. Anyone interested in the topic should have a look at the article of Varian (2006). Someone willing to understand more deeply this literature and its history should read the article of Gardes and Garrouste (2006). We will develop further on the relation between the proposition of Bernheim and Rangel (2009) and revealed preferences in section 2

served choice. That is, the “environment” of the choice may affect it, by changing its perceived characteristics, but in reality it does not affect the alternative itself and the welfare derived from it.

However, if the underlying idea is the same, the use of frames is very different. Rubinstein and Salant (2012) use them to correct biases of individuals. That is, they use the knowledge we have of the context of the decision to look for a distortion function that goes from the underlying welfare order to the observed choices. It means that an observed choice order does not necessarily directly reflects the welfare order. The constraint of the model is that using the welfare ordering and the distortion function together should yield to the observed choice. Obviously, to yield to correct recommendations, this approach relies on the choice of the “good” distortion by the modeler. It also means that for each situation, the model might be different. On the other hand, if the model is properly built, welfare inferences are correct. On the contrary, Bernheim and Rangel (2009) try to build a model that is very general and could be applied in a lot of situations with the same hypotheses. This proposition and its consequences are explained in the next subsection (1.2), but before that, let us review some other attempts.

Apesteguia and Ballester (forthcoming) suggest a measure of rationality and welfare, called the *swap index*. This measure is a way to evaluate both the inconsistencies and the welfare of agents. The aim is to choose the closest consistent preference order to (empirically) observed choices. To assess the closeness of a preference order, they look at the number of swapping of alternatives necessary to get the preference order from the observed choices. Think of the following example: the welfare order is $A > B > C$, but observed choices are $A > C > B$, then $C$ and $B$ have to be swapped to get the welfare order. To be applicable however, it needs a complete observed order, as when two elements are not comparable, swapping is not possible. They finally show that their model encompasses others and in particular that the rationality measures of Varian (1990), the *goodness-of-fit* measure, and Houtman and Maks (1985), the *HM-index* fall in their setting

Another axiomatization of welfare inference has been proposed by Chambers and Hayashi (2012). Broadly speaking, they introduce an individual welfare functional, which is a function from each situation choice distribution (that is, the distribution of choice on a set: if the set contains two elements $\{A, B\}$, each one chosen half of the time, then the distribution is the uniform one), to a complete and transitive binary relation on all alternatives. Then they provide some axioms to characterize this individual welfare functional.

Finally, the *transitive core* has been introduced by Nishimura (2014). In a sense, it is close to what will be done here, as it uses preference relations. He suggests to find the set where preferences are transitive and axiomatize it. The insight is that on the transitive part of the relations, sarp is verified and recommendations are possible without relying on any arbitrary decision of the modeler, similar in this sense to Bernheim and Rangel (2009), with the big caveat that observed choices must be the welfare maximizing ones. He also provides an algorithm to get this transitive core and equivalence to some other models, like the $\beta\delta$-model of time inconsistency and prospect and regret theory.

## 1.2 The Strict Unambiguous Choice Relation

The Strict ucr has been introduced by Bernheim and Rangel (2009), with two main purposes:

- Build sounder theoretical foundations welfare inferences in behavioral economics;
- Base welfare on observed choices, as the main observation available to economists (and thus following the revealed preference literature).

The core idea of the Strict ucr is simple:

> $x$ is strictly unambiguously chosen over $y$ if and only if $y$ is never chosen when $x$ is available, and $x$ is chosen at least once.[2]

It should be noted that Green and Hojman (2007) reached the same preference relation, using completely different premises. In their model, individuals are multiple selves with conflicting preference relations. They showed that under certain axioms, this model leads to Strict ucr.

For readers not familiar with this literature, it is useful to compare it to the more traditional revealed preference relation:

> $x$ is revealed preferred to $y$ if and only if $x$ is chosen when both $x$ and $y$ are available and $y$ is not chosen.

The fundamental difference between the two is the following: with Strict ucr, all observations are considered to infer preferences, whereas with revealed preferences, each observation is taken independently. The two approaches are related and we will elaborate the link between revealed preferences and Strict ucr in section 2.4.

One feature of Strict ucr that drew attention is its acyclicity under some hypothesis. It is in theory a very appealing result. It means that the preference relation obtained is transitive, and that there is at least a partial order over observations. This result is from the early literature of revealed preferences and in particular Arrow (1959). The hypothesis, which we call *complete data*, is the following:

**Definition 1** (Complete data). *A dataset of observation is said to be* complete *if and only if for each possible subset of the grand set of alternatives $\mathcal{X}$, we observe at least one choice.*

This hypothesis is quite restrictive, indeed, if $n$ is the number of alternatives in $\mathcal{X}$, then we need at least $2^n$ observation[3]. This is a very demanding assumption in general and in particular in our setting of consumption data. Indeed, there is only one paper as far as we know that have complete data for its application, Manzini and Mariotti (2010)'s "Revealed preferences and boundedly rational choice procedures". In general preferences should exhibit some cycles and thus be intransitive. The situation may emerge pretty easily, as example 1 of Bernheim and Rangel (2009) shows. One question here is to know whether it will arise or not.

[2] It must be noted that the original definition uses the notion of frames as introduced by Tversky and Kahneman (1981), but for the sake of simplicity, and because we do not need it here, we dropped it.

[3] In fact a bit less, as we have no choice from the $\emptyset$, and choices from singletons is trivial, thus we are left with $2^n - n - 1$.

### 1.3 Summary of results

Detailed results will be presented in section 3, but the salient ones are:

- Most households (99.7%) are not utility maximizers according to revealed preferences, whereas most are (90%) according to Strict UCR. The cyclicity of some households is a direct result of the incompleteness of our dataset.
- The empirical preferences are almost complete, with an average number of relation per bundle above 46, whereas here the theoretical maximum for a rational household is 47. There is caveat however, as our household are not necessarily "rational".
- We introduced a new relation, Strict UCR+, which guarantees acyclicity and is for us a quite natural extension of Strict UCR.
- Strict UCR+ is almost complete and we can close it transitively. It means that we are able to perform prediction and use it to compute the Selten's score. That show another light on the tightness of the predictions and thus on the usefulness of the theory.
- Predictions made with Strict UCR+ are tight, which means that the theory is both useful and predictive.
- Finally, from a theoretical point of view, the set of preference relation obtained by Strict UCR is the same as the set of preference relations obtained through revealed preferences restricted to where the Weak Axiom of Revealed Preferences (WARP) is verified, in the case of choice function (that is, choices are single peaked).

To show those different results, the paper is divided in three sections. In section 2, we introduce Strict UCR and Strict UCR+. In section 3, we present the results of the application of Strict UCR and Strict UCR+ to consumption data. In section 4, we present results of an application of Strict UCR+: predictions and Selten's score.

# 2 Strict UCR and Strict UCR+

## 2.1 Strict UCR

### 2.1.1 Definition

In the introduction, we defined Strict UCR as:

> x is strictly unambiguously chosen over y if whenever x and y are both available, y is never chosen and x is chosen at least once.

It is time to define this preference relation more formally, so that we properly and unequivocally apply it. Denote C a choice correspondence (that is, a function that associates to a subset A of $\mathcal{X}$ a subset of A). Formally, noting $P^*$ the Strict UCR:

$$xP^*y \Leftrightarrow \begin{cases} \exists A \subset \mathcal{X}, x, y \in A, x \in C(A) \\ \forall A \subset \mathcal{X}, x, y \in A \Rightarrow y \notin C(A) \end{cases}$$

This definition uses choice correspondence and we should translate it for an application on consumption data. In order to do so, we need to make a detour through revealed preferences and single peaked choices (choice functions).

### 2.1.2 Revealed Preferences

Throughout this paper, a preference relation will be noted R when weak, and P when strict. Formally:

- $xRy$ means that alternative x is *weakly preferred* to y (leaving the possibility for a consistent agent to exhibit the preference relation $yRx$).
- $xPy$ means that x is *strictly preferred* to y, that is $xRy$ and not $yRx$.
- Finally, if $xRy$ and $yRx$, we say that x is *indi erent* to y, and note it $xIy$.

These are general definitions of preference relations. In the specific situation of consumption data, the primitive is a couple $(x, p_x)$, where x is a bundle (=vector) of goods and $p_x$ is the associated vector of prices. In that context, we say that:

- x is (weakly) directly revealed preferred to y, denoted $xR^0y$:

  $$xR^0y \Leftrightarrow p_x x \geq p_x y$$

  An external observer observes once that x is chosen when y is available, but it does not rules out the converse (y is chosen when x is available).

- x is (strictly) directly revealed preferred to y, denoted $xP^0y$:

  $$xP^0y \Leftrightarrow p_x x > p_x y$$

  An external observer observes once that x is chosen when y is available, and it rules out the converse. If the converse is ever observed, this leads to an intransitive behavior, and the agent could be money pumped. Obviously, $xP^0y$ implies $xR^0y$.

- x is (weakly) revealed preferred to y, denoted $xR^Cy$:

  $$xR^Cy \Leftrightarrow \exists (x_i)_{i=1,\ldots,n}\ xR^0x_1, x_1R^0x_2, \ldots x_nR^0y$$

  The preference relation is not necessarily directly observed, but can be deduced from a sequence of observation. We have that $xR^0y \Rightarrow xR^Cy$.

- x is (strictly) revealed preferred to y, denoted $xP^Cy$:

  $$xP^Cy \Leftrightarrow \exists (x_i)_{i=1,\ldots,n}\ xR^0x_1, x_1R^0x_2, \ldots, x_nR^0y \text{ and one relation is } P^0$$

  It is the transitive closure of the relation of strict direct revelation. The main point here is that only one relation in the sequence needs to be strict to get the strict preference relation. $xP^0y \Rightarrow xP^Cy$ and $xP^Cy \Rightarrow xR^Cy$.

These relations are illustrated in figure 2.1.


Figure 2.1: Example of weak and strict revealed preferences. The slopes of the budget lines are the ratio of prices. Observations 1 and 4 share the same revenue and (ratio of) prices, a situation that never occurs in our data.
Here $AP^0B$, $AR^0D$, $BP^0C$, $BP^0D$, $CP^0D$, $DR^0B$ and $DP^0B$. The observations $BP^0D$ and $DP^0B$ are inconsistent (they do not verify warp). By transitive closure, we have that $AP^CC$, whereas A and C are in general not directly comparable. $DP^CC$ is also true, which is inconsistent with $CP^0D$.

### 2.1.3 Strict UCR applied to consumption data

The intuition of the definition in the case of consumption data is the following. Consider two observations $(x, p_x)$ and $(y, p_y)$. $y$ is *available* in the situation where $x$ is chosen if and only if $p_x y \leq p_x x$. It means that Strict ucr in that context can be written as:

$$xP^*y \Leftrightarrow \begin{cases} p_x x \geq p_x y \\ \neg p_y y \geq p_y x \end{cases} \Leftrightarrow \begin{cases} xR^0y \\ \neg yR^0x \end{cases} \Leftrightarrow \begin{cases} p_x x \geq p_x y \\ p_y y < p_y x \end{cases}$$

Where $xP^*y$ is used to say that $x$ is strictly unambiguously chosen over $y$. Despite the similarity, it is different from the condition $xPy$, because it involves two different observations. If prices are such that $p_x = p_y$, then $xP^*y \Leftrightarrow xP^0y$.

In words, the condition above means that $y$ should be available when $x$ is chosen, whereas $x$ should not be available when $y$ is chosen. If $y$ is not available, then $x$ and $y$ are incomparable. Whereas if $x$ is available when $y$ is chosen, the choice is ambiguous and thus does not generate a preference relation.

## 2.2 Caveat of Strict UCR

Strict ucr, as a leading proposition, has attracted various critiques. One set of critiques came from Salant and Rubinstein (2008). Their main concern is that Strict ucr does not yield to the real welfare order. Indeed, by using observed choices as the primitive for the preference relation, biases are not corrected. If the observed choice is not the welfare maximizing one because of a bias of some kind, then the preference (and welfare) order assessed will be wrong too. Partly to address this concern, the full definition of Strict ucr as introduced by Bernheim and Rangel (2009) use an ancillary condition in addition to the observed choice. This ancillary condition is a variable that allows to take into account the "environment" of the choice, in addition to the choice itself. In particular, they compare only choice performed with the same environment and not choices performed in different environment, so that biases should be at least constant across the preference relation obtained. This does not mean that the bias vanishes, so that it by no means guarantee that the preference order obtained is the real welfare maximizing one[4].

We will not address this concern in the present work. We chose to ignore the ancillary condition for several reasons. The first one is that all choices are choices from grocery stores, so that the situation of choice is more or less the same. Obviously it is not exactly the same. Indeed the entity here is the household and their might be several members going shopping. The second reason is a practical one. We do not have data about the environment of choice. Finally, keeping all observations makes the test harder for Strict ucr. Indeed, if large biases are at hand, they should generate at lot of "irrational" choices and thus make it harder to get a transitive preference relation.

Another concern pointed out by Rubinstein and Salant (2012) among others is the tightness of Strict ucr. Indeed, taking all observed choices together to assess preference relations may be too demanding. In such a case, Strict ucr would lead to a preference relation that looks like a welfare maximizing one, but would be too empty to be really usable. It would be theoretically appealing but descriptively useless and would not provide meaningful predictions. We will test this concern in this work, in two steps. First,

[4] if this notion is meaningful of course.

we will try to assess the preference relations that come out of the data and see whether it is complete or not. Second, we will try to make predictions, following the methodology of Selten (1998) and Beatty and Crawford (2011) among others.

A final caveat of Strict ucr comes from one of its assumption. In order to prove that Strict ucr is acyclic, $P^*$ need complete data. That is, for all subsets of the grand subset of alternatives, a choice must be observed. This hypothesis is very demanding in empirical settings. The question then is how dropping this assumption will affect the results. Acyclicity is not guaranteed anymore, at what cost. That is, does Strict ucr perform significantly better than revealed preferences with that respect? To deal with the incompleteness of Strict ucr we introduce a new relation, called Strict ucr+, that guarantees acyclicity.

## 2.3 Strict UCR+

The objective of this section is to introduce a refinement to Strict ucr, if possible leading to the same preferences when Strict ucr is acyclic (and in particular on complete data). The other purpose of the refinement is to guarantee acyclicity, whether the data is complete or not, thus answering the last remark of the previous section.

For this purpose, we introduce a new relation, called the *strict unambiguous choice relation+* (Strict ucr+ hereafter), which is defined as:

**Definition 2** (Strict ucr+).

$$xP^{**}y \Leftrightarrow \begin{cases} xP^*y \\ \nexists n \geq 1, (x_i)_{1 \leq i \leq n} \text{ such that } yP^*x_1P^* \dots P^*x_nP^*x \end{cases}$$

*In words,* $P^{**}$ *is* $P^*$ *restricted to transitive relations.*

This definition leads obviously to the same result as Strict ucr when the latter is acyclic. In particular, both relations coincides when the data is complete. When it is not the case, it guarantees acyclicity by removing cycles of $P^*$. It should be noted that we have $n \geq 1$: no cycles of size 2 can emerge definition, because in that case the choice would be ambiguous and we would not get any $P^*$ relation. One can show (proof is in appendix 6.1), that it is equivalent to removing from revealed preferences, first cycles of size 2. Then from the remaining relations, removing all cycles. In addition to this, to perform out of sample predictions, we transitively close the relations we have obtained.

## 2.4 Link with the revealed preferences axioms

In this part, we show that applying Strict ucr in the case of choice functions, where choices are single valued is equivalent to imposing the weak axiom of revealed preferences to revealed preferences. More precisely, the set of Strict ucr relations is the same set as the set of revealed preferences relations restricted to relations that verifies the weak axiom of revealed preferences.

Remember that warp is the following (Varian 2006):

$$aP^0b \text{ and not}(bR^0a)$$

The last part can be rewritten in terms of choices from menus:

$$\text{not}(bP^0a) \Leftrightarrow \forall A \text{ s.t. } a, b \in A, b \neq C(A)$$

That is, a must be strictly revealed preferred to b, whereas b must never be weakly (or strictly, with choice functions it is the same), revealed preferred to a. This is a reminder of the Strict ucr relations, which indeed in that situation writes:

$$aP^*b \Leftrightarrow \begin{cases} \exists A \text{ s.t. } a, b \in A, a = C(A) \\ \forall A \text{ s.t. } a, b \in A, b \neq C(A) \end{cases}$$

The first line is obviously $aP^0b$ whereas the second is the warp restriction. It means that $P^*$ is significantly different from imposing warp on revealed preferences only in the case of choice correspondence. It is an interesting result if you remember the original purpose of Bernheim and Rangel's paper: building choice theoretic foundations to behavioral (welfare) economics. Indeed, the most common tool in there is revealed preferences where choices are usually unitary because of the observation by observation nature of the approach. warp is not a very good concept however, as it leaves cycles. The good concept to get rid of non-transitivity is sarp. The latter however needs the transitive closure to be applicable, so in that sense it is not only observationally based.

# 3 Application

In order to address concerns raised in the previous section, we applied the Strict ucr to consumption data, more precisely to package grocery store data, from Aguiar and Hurst (2007). The method used to build the prices and bundles and thus the preference information follow the standards of revealed preferences testing, with all its pros and cons (see Crawford and De Rock (2014) for a review of methods). A more precise description of the data is available in the next subsection, followed by the results obtained and a comparison with some rationality measures.

## 3.1 Data

The data has been collected by the marketing firm ACNielsen. It has been originally used in economics by Aguiar and Hurst (2007) in their article "Life-Cycle Prices and Production". We use the same method as Dean and Martin (forthcoming) to process the raw data, in particular, we use the same goods categories and prices. The full description of the original dataset, and in particular its demographic characteristics, can be found in the latter, whereas a description of the subsample used here is available in the former, as well as in appendix 6.2.

Here, we restrict our attention to a subsample of 397 households, for which a purchase was made in all 2-weeks time periods between February 1993 and 1995 (the original data is made of around 2,100 households). This sample is a balanced panel over the period, thus avoiding the problems of repeated cross-sections in terms of comparability. ACNielsen is a marketing firm, and they attached great importance to the reliability of their data. Households document the upc (universal product code) of the products they buy at home, regrouping all shops they have been in. Thanks to the upc, prices and characteristics of

the goods at the moment of the purchase are recovered. The procedure is voluntary, so that reports might not be truthful. Respondents have been paid, but they may still lie. Even though neither representativity nor reliability really matters here, as we want first to test the Strict ucr on real data and not to claim anything on the "rationality" of the households.

For comparability reasons, goods are aggregated in 36 categories, the list is in appendix ??. Products that do not fall in this classification are excluded. The two years time length leads to 48 two-weeks periods. The characteristics of the data are the following: 397 households, with an average of 18.1 purchases, 3.5 store trips, and $20.4 in expenditure per period. The demography of this sample is available in appendix 6.2.

The aggregation over time is to alleviate concerns about storage of some items (say, pastas or rice). However, aggregating over long periods reduce the number of observations and the probability of violations. That is the main reason why we choose to aggregate over two weeks and not a month or longer[5]. Price indices are constructed for this periods, using the classical method in the literature of "Stone" prices:

$$P_{Jt} = \sum_{i \in J} w_{it} p_{it}$$

where $P_{Jt}$ is the price index for good category J in the period t, $w_{it}$ is the budget share for UPC code i in the period t, and $p_{it}$ is the mean price for UPC code i in period t. Dean and Martin (forthcoming) examines also other indices and do not find significant differences, thus we keep that one and do not investigate more.

Here of course, not all items a household can buy are represented, the average sum used per period is here to testify that. For instance, leisures are not in our sample. To get acyclicity from utility maximization on a restricted sample of all possible purchases, the assumption of additive separability between the goods considered and the others is required. It is quite strong, but also standard in revealed preferences testing. It is also assumed that all households faces the same prices in a given period. This assumption is necessary, because it might happen that a household does not buy a category of good in a given period, thus revealing no prices for that category and hampering revealed preferences testing. It could be possible to keep only households that have bought all types in all periods, but it would restrict too much the sample[6]. Of course, in reality, prices vary between stores and in a given period, we thus introduce an error, but it is standard and the error is likely to be small (goods prices in a two-week time period will not double). Another concern with that method is linked to what prices really face the household: buying a ready made salad and buying all ingredients and making it arguably lead to a similar consumption, but prices might be very different, and monetary prices might not reflect that difference. In fact, a hand made salad has also a cost of time, which the ready made salad does not have. The data does not allow us to take into account that but we do not believe it is a big problem as most of the food purchased in the USA in general and in our sample in particular is ready made.

[5] Analyses have also been performed on a monthly time period and with only 3 aggregate goods. The results are unsurprising: less violations, with some remaining. We choose that case as it is more stringent and thus more interesting for our proof of concept.

[6] Dean and Martin (forthcoming) found that it would reduce the sample size by 85% for the monthly sample; It would be even more stringent in our case.

| | Average | Median | Std Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Cycles of RP | 661 | 14 | 1196 | 1 | $150,027$ |
| Cycles of Strict UCR | 2 | 1 | 1.4 | 1 | 18 |

Table 3.1: Statistical description of the number of cycles per household, conditional on being a cyclic agent.

As budget for households, we used total expenditures of that period. This introduce the well-known problem of revealed preference testing, that was pointed out by Varian (1982): budget lines might not intersect a lot across periods, thus revealed preference testing is vacuous. We will come back to that point in section 4, where we show how tight the prediction of the models are.

The last important assumption made for the empirical testing is the stability of preferences across time. This assumption is needed to compare across periods. If preferences were to change, then having a violations of revealed preferences would only mean that preferences have changed and would not be informative *per se*. This assumption is tested in appendix 6.3, for robustness check and is essentially verified.

One important feature of the data is that two bundles never comes twice for the same household. Prices are always different in two different periods. This means that preference relations are always strict, as inequalities are strict. It means that we can use the Arrow formulation of revealed preferences, and use the choice function definition. Thus in our example, Strict UCR is equivalent to restricting revealed preferences to where WARP is verified.

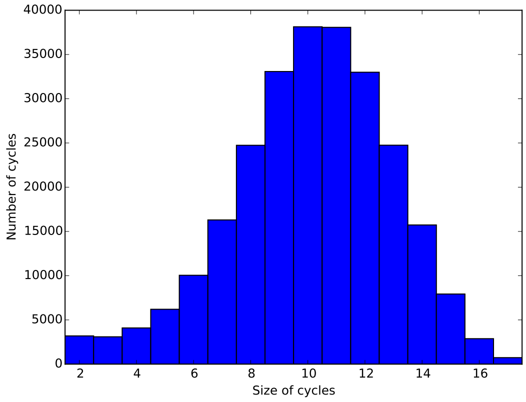
## 3.2 Cyclicity

The first question raised is whether Strict UCR is cyclic, as the incompleteness of the data may lead to. We show here it is, but much less than what revealed preferences are. Moreover, it is less cyclic in three manners:

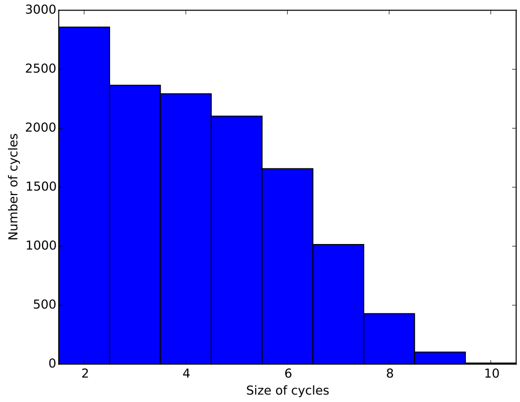
- There are less households violating GARP (= cyclic): 396 violate GARP with revealed preferences, compared to 40 with Strict UCR and 0 with Strict UCR+;
- Each cyclic household exhibit less cycles on average for Strict UCR compared to revealed preferences (median number of cycles of 12 and 1, respectively, see table 3.1 for more precise statistics);
- Cycles are shorter, on average.

The first point comes directly from the fact that Strict UCR is more restrictive than revealed preferences, and in particular that it "removes" cycles of size 2. By not giving birth to those cycles, it effectively removes most of the bigger ones.

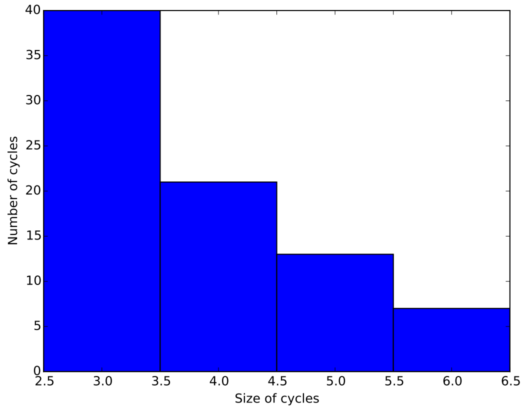
The second is illustrated on table 3.1, which shows there are much less cycles for cyclic households, as the median goes from 12 to 1. Average is not a very good statistics because the distribution of numbers of cycles by households has a very fat tail, with some households with a very high number of cycles. Those households have some small cycles connected, that somehow generates a base for other cycles. If you connect a lot of cycles




(a) Revealed Preferences




(b) 95% of the revealed preferences sample, that is, removing households with the higher number of cycles.




(c) Strict ucr

Figure 3.1: Histograms of the number of cycles, scales are very different.

of size 2, it is possible to generate much longer cycles and a lot of cycles of intermediate size. This fact is perfectly illustrated by the three histograms of figure 3.1 (be careful of the scale on the left). Strict ucr has a very small number of cycles. They are also shorter, with a size going from 3 to 7. On the other hand, revealed preferences exhibit much more and longer cycles, with a size going from 2 to 17. These results show a bad light on revealed preferences. However, this higher number of cycles comes only from a small number of household. Looking at figure 3.1b with the 95% less cyclic households show it. Now there are a lot of small cycles, but longer ones have subsided.

Another way to look at that is table 3.2, which shows the absolute number of relation removed in a Strict ucr and Strict ucr+ compared to revealed preferences. These number are very small, especially if you remember that the maximum number of cycles is 150,027. It is possible to get rid of all of them while removing only 53 relations. It is even possible to do it while removing less relations, especially since here we constrained ourselves to removing all the relations of a cycle, which is not necessary if you just want to have transitive preferences. Indeed, in any cycle, you only need to remove relation to

| | Average | Median | Std deviation | Maximum |
|---|---|---|---|---|
| Strict ucr | 16.1 | 14 | 6.9 | 50 |
| Strict ucr+ | 16.6 | 14 | 7.2 | 53 |

Table 3.2: Statistical characteristics of the number of relations to remove to rationalize the preferences of an agent. Strict ucr removes slightly less but leaves some cycles.

| | Average | Std Deviation | Minimum | Maximum |
|---|---|---|---|---|
| Revealed preferences | 46.80 | 0.22 | 45.75 | 47.54 |
| Strict ucr | 46.12 | 0.25 | 44.83 | 46.75 |
| Strict ucr+ | 46.11 | 0.26 | 44.46 | 46.75 |
| Closed Strict ucr+ | 46.29 | 0.21 | 44.96 | 46.92 |

Table 3.3: Average number of relation per bundle, the theoretical maximum is 47 (each bundle of an agent should compare to all others). Closed indicates that it is the transitive closure.

restore transitivity. By properly choosing the relations removed, it is possible at a smaller cost in terms of preference relations. We have not chosen this path here however, as we do not want to choose the relation to remove in a given cycle and we have no rule for that. An example of such an approach is provided by the *minimum cost index* of Dean and Martin (forthcoming).

## 3.3 Completeness

The second concern regarding Strict ucr that came out is on the tightness of the relation: it might be too strong so that the preference relation would be empty. The preferences are transitive (which is mostly the case), but at the cost of having no relation left, which hampers its descriptive power. This concern should be assuage by the results shown on table 3.3. Whereas the theoretical maximum for a "rational" household is to have 47 relations per bundle (or equivalently 1128 relations in total), the average over all households is 46.12. There is a caveat to this result however: 40 households still exhibit cycles.

It leads to Strict ucr+. With it, all households exhibit transitive preference and the average number of relations is 46.11, with now the possibility to close transitively the preference relation. Doing so, the average number of relations is 46.29, which is higher than Strict ucr. It is however lower than what revealed preferences lead to, with the average of 46.80, with the big caveat that most households are cyclic there. And here the average does not hide big differences, as a look at the standard deviation and the minimum and maximum in the table shows. Another way to look at that fact is to look at the empirical CDF of the number of relations per households in figure 3.2. Revealed preferences have a higher number of relations. All distributions are very narrow, and removing all cycles is not very costly in the end (removing all cycles for revealed preferences costs much more in terms of relations, results available in appendix 6.4).

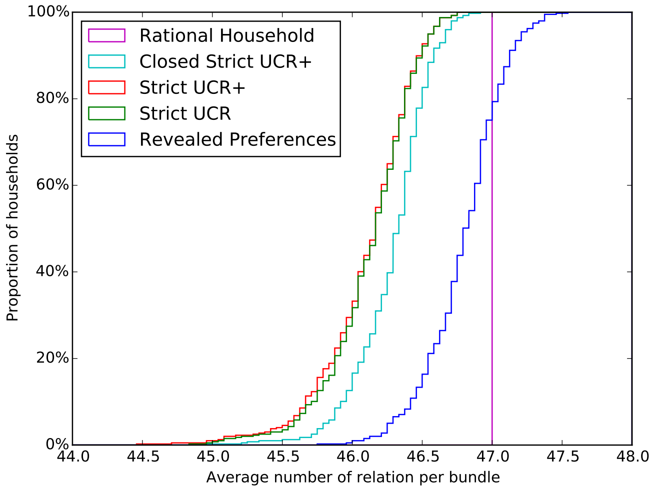



Figure 3.2: CDF of the distribution of the households average number of relations per bundle, for the relations introduced in this paper. Closed is used to denote the transitive closure. Rational shows the theoretical number of a rational household.

# 4 Application: Out of sample prediction

## 4.1 The Selten's score

The aim of this part is twofold:

- Provide an illustration on how the theory developed until now can be applied;
- Assess the explanatory power of the theory, with in mind whether the theory is tight or not.

With those two purposes in mind, we follow the idea first introduced by Bronars (1987) to test for the power of revealed preferences tests. It has been formalized and axiomatized by Selten (1998) and used, among others, by Beatty and Crawford (2011) and Dean and Martin (forthcoming). In this setting, theories can be divided in three categories:

- A theory with vacuous predictions: all observations are admissible. The theory might explains quite nicely things, but do not restrict observations and is thus uninformative.
- A theory wih tight predictions never verified in empirical settings. This theory lacks descriptive power.
- A theory with tight predictions verified in empirical settings. This theory is the best we can achieve.

It turns out that this idea may be captured by a simple tool, called the Selten's score, as axiomatized by Selten (1998). It is made up of two ingredients:

- An index telling whether the real observation verifies the theory, denoted $r$ from now on. It aims at testing the theory on real data;
- Another telling whether a random observation verifies the theory or not, denoted $a$ hereafter and called the *pass rate*. It verifies whether the theory is constraining or not.

The Selten's score $S$ is a combination of both:

$$S = r - a$$

Here, because on how we built our indices, we will focus on the $a$ index more than on the overall Selten's score: we are interested on whether the theory is tight or not, and not whether it fits real observation, something we have already discussed in the previous sections. Moreover, we use median budget at the household level with no corresponding observation. We introduce the whole method because we have also computed the Selten's score using the real budget (results in appendix 6.6).

To test for the tightness of the theory, we work at the household level (even though we will show aggregate results):

1. Drop one observation (to fix ideas, the last period observation);
2. Apply Strict ucr+ to the remaining observation and close transitively the preference relation obtained;
3. (supplementary) Look if the real observed bundle would pass or fail garp. This give us $r = 0$ (fail) or $r = 1$ (pass).
4. Draw random bundles with the median budget and the dropped period prices (here the last period prices) to map the price hyperplane and look at how many of them would pass or fail garp. We get a number $a \in [0, 1]$ which is the probability that a random bundle passes garp. If $a = 1$, all passes, whereas if $a = 0$, no one pass.
5. (supplementary) Compute $S = r - a$.

The precise steps are detailed in appendix 6.5, and in particular the way we have built random bundles. It is an important technical detail, because for each period, most goods are not chosen, and thus we get corner solutions. It means that using a uniform distribution on the simplex is not appropriate to map the budget shares, and lead to misleading numbers.

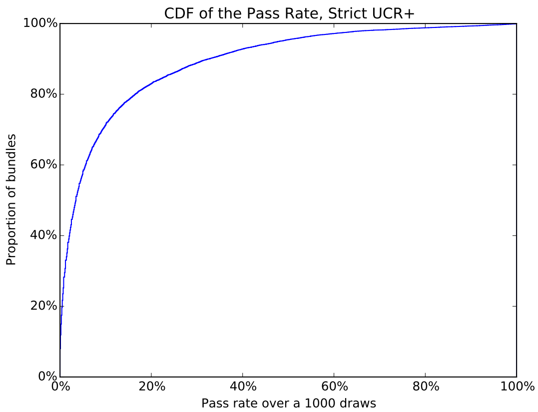
## 4.2 Results

The steps explained above have been performed for all observations, that is, all households at all periods and we show aggregate results in table 4.1 as well as some more detailed ones in figure 4.1a. As can be seen, the average pass rate is around 0.11: the theory is tight, whereas the Selten's score is around 0.34, which strengthen that impression. The theory is restrictive, maybe too much.

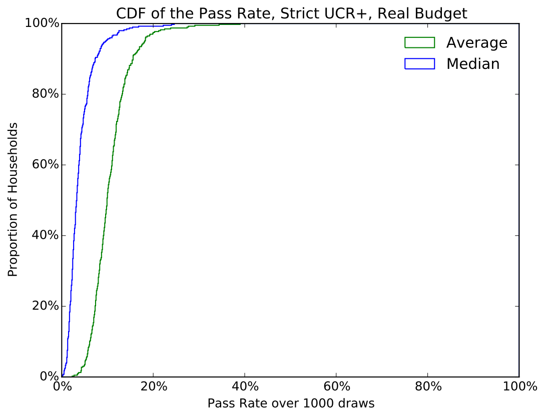
A more interesting result is to see where the variability occurs. Indeed, two extreme possibilities are:

| | Aggregated | | Median by hh | | Average by hh | |
|---|---|---|---|---|---|---|
| | Pass Rate | Selten | Pass Rate | Selten | Pass Rate | Selten |
| Average | 0.11 | 0.35 | 0.04 | 0.02 | 0.11 | 0.35 |
| Median | 0.04 | 0.00 | 0.03 | 0.00 | 0.10 | 0.35 |
| Standard Deviation | 0.17 | 0.52 | 0.03 | 0.07 | 0.04 | 0.04 |
| Minimum | 0 | -1 | 0.00 | -0.01 | 0.02 | 0.09 |
| Maximum | 1 | 1 | 0.25 | 0.46 | 0.39 | 0.44 |

Table 4.1: Statistics of pass rate and Selten's score, on a 1000 random draws for each observations. The median has been first calculated by households, then over households.




(a) Area where garp passes, all observations are used.




(b) Area where garp passes, first aggregated at a household level.

Figure 4.1: Empirical cumulative distribution of the pass rate. The numbers have been assessed for each observation by a 1000 random draws.

1. Each household has a very narrow range of pass rate: at the extreme, one pass rate, identical for all observations;

2. All the variability is at the household level: each one has a very wide range of pass rates, and the aggregate is only the reflexion of that variability.

A first answer to that question is provided when we compare figures 4.1a and 4.1b. In the former one, the CDF looks like a $1 - \exp x$ function: a lot of pass rates are near zero, whereas very few are close to one. It means that for most observations, the pass rate is close to nil. On the other hand, when looking at the numbers aggregated at the household level (figure 4.1b), it can be seen that the distribution of the median and the average per household differ. It hints towards the following distribution of the pass rate for each household: most of them close to zero, with a small fraction significantly higher than zero. This drives the average up compared to the median.

This interpretation shore up by looking at one household, as in figure 4.2. The figure shows that most of the pass rates are between 0 and 10%, with some of them really higher, up to 50%. Both observations lead to think that most of the variability is inside households, closer to explanation 2, rather than explanation 1.

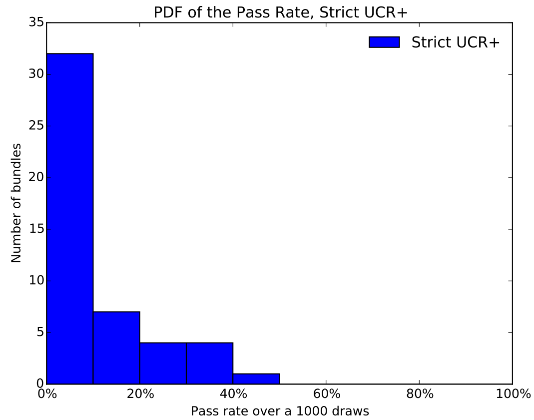



Figure 4.2: Histogram of the pass rate for one household, build over the 48 observations.

The results show that when we use a budget made such that it crosses well other budgets set (by using the median budget for each household), the theory is restrictive. This tightness is well dispersed over all households and not concentrated on some. On the other hand, if we use the real budget of the observation, results are less rosy: the pass rate is higher (see 6.6). It means that the choice of the budget line is important to study the tightness of predictions.

# 5 Conclusion

In this work, we provided the first application of the *strict unambiguous choice relation*. In particular, we applied it to consumption data, a purpose that might not be the first one that come to mind but that uses one of the most widely available type of data. We compared the results obtained to those of revealed preferences and shows it is better, in the sense that it is less cyclic but still keeps a lot of preference relations. The cyclicity of Strict ucr comes from the incomplete nature of the data that we have. To alleviate the problem, we introduce a new preference relation, Strict ucr+, that is equivalent to Strict ucr when the former is acyclic and guarantees acyclicity otherwise. This is done while keeping a relatively complete preference relation, which was one concerned about Strict ucr. Strict ucr+ allowed us to compute Selten's score and to look at the tightness of predictions, which quite heavily depends on the choice of the budget, but can be tight with Strict ucr+.

## References

Afriat, Sidney N (1967). "The construction of utility functions from expenditure data". In: *International economic review* 8.1, pp. 67–77.

Aguiar, Mark and Erik Hurst (2007). "Life-Cycle Prices and Production". In: *The American Economic Review*, pp. 1533–1559.

Apesteguia, Jose and Miguel Angel Ballester (forthcoming). "A Measure of Rationality and Welfare". In: *Journal of Political Economy*.

Arrow, Kenneth J. (1959). "Rational Choice Functions and Orderings". In: *Economica* 26.102, p. 121. issn: 00130427. doi: 10.2307/2550390. (Visited on 04/13/2015).

Beatty, Timothy K. M. and Ian A. Crawford (2011). "How Demanding is the Revealed Preference Approach to Demand?" In: *The American Economic Review* 101.6, pp. 2782–2795. issn: 0002-8282.

Bernheim, B Douglas and Antonio Rangel (2009). "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics". In: *The Quarterly Journal of Economics* 124.1, pp. 51–104.

Bronars, Stephen G. (1987). "The Power of Nonparametric Tests of Preference Maximization". In: *Econometrica* 55.3, pp. 693–698.

Chambers, Christopher P and Takashi Hayashi (2012). "Choice and Individual Welfare". In: *Journal of economic theory* 147.5, pp. 1818–1849.

Clippel, Geoffroy de and Kareen Rozen (2012). "Bounded Rationality and Limited Datasets: Testable Implications, Identifiability, and Out-of-sample Prediction". In:

Crawford, Ian and Bram De Rock (2014). "Empirical Revealed Preference". In: *Annu. Rev. Econ.* 6.1, pp. 503–524.

Dean, Mark and Daniel Martin (forthcoming). "Measuring Rational with the Minimum Cost of Revealed Preference Violations". In:

Gardes, F. and P. Garrouste (2006). "Jean Ville's Contribution to the Integrability Debate: The Mystery of a Lost Theorem". In: *History of Political Economy* 38 (Suppl 1), pp. 86–105. issn: 0018-2702, 1527-1919. doi: 10.1215/00182702-2005-018.

Green, Jerry R. and Daniel A. Hojman (2007). "Choice, rationality and welfare measurement". In:

Houtman, Martijn and J Maks (1985). "Determining all Maximal Data Subsets Consistent with Revealed Preference". In: *Kwantitatieve methoden* 19, pp. 89–104.

Koo, Anthony YC (1963). "An Empirical Test of Revealed Preference Theory". In: *Econometrica: Journal of the Econometric Society*, pp. 646–664.

Manzini, Paola and Marco Mariotti (2010). "Revealed preferences and boundedly rational choice procedures: an experiment". In: *University of St. Andrews and IZA. Working Paper. Make Us Smart., ed. Gerd Gigerenzer, Peter M. Todd and the ABC Research Group*, pp. 141–167. url: http://www.st-andrews.ac.uk/~pm210/similEXPERIMENT.pdf (visited on 06/05/2015).

Nishimura, Hiroki (2014). "The Transitive Core: Inference of Welfare from Nontransitive Preference Relations". In:

Rubinstein, Ariel and Yuval Salant (2012). "Eliciting Welfare Preferences from Behavioural Data Sets". In: *The Review of Economic Studies* 79.1, pp. 375–387.

Salant, Yuval and Ariel Rubinstein (2008). "(A, f): Choice with Frames". In: *The Review of Economic Studies* 75.4, pp. 1287–1296.

Samuelson, Paul A (1938). "A Note on the Pure Theory of Consumer's behaviour". In: *Economica*, pp. 61–71.
Selten, Reinhard (1998). "Axiomatic Characterization of the Quadratic Scoring Rule". In: *Experimental Economics* 1.1, pp. 43–62.
Tversky, Amos and Daniel Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases". In: *science* 185.4157, pp. 1124–1131.
— (1981). "The Framing of Decisions and the Psychology of Choice". In: *Science* 211.4481, pp. 453–458.
Varian, Hal R (1982). "The Nonparametric Approach to Demand Analysis". In: *Econometrica: Journal of the Econometric Society*, pp. 945–973.
— (1990). "Goodness-of-fit in Optimizing Models". In: *Journal of Econometrics* 46.1, pp. 125–140.
— (2006). "Revealed Preference". In: *Samuelsonian Economics and the Twenty-first Century*, pp. 99–115.

# 6 Appendices

## 6.1 Appendix: Equivalent of Strict ucr+ with removing cycles of size 2 of revealed preferences, then all cycles left

These two steps comes from the two di�erent step of Strict ucr+. In the first step, we apply Strict ucr and in the second we remove all cycles. So the first step of the proof is to show the equivalence of Strict ucr with removing cycles of size 2 when we are in a choice function context. The second is obvious once the first is proven.

**Proof:** We will demonstrate that cycles to be removed of revealed preferences with these assumptions are of length 2, that is, contains only two di�erent alternatives. In the case of cycles of length 2 for revealed preferences, $P^*$ will not generate binary relations, because it means that the choice is ambiguous: there exists two subset $A, B \subseteq X$, such that $a, b \in A \cap B$, with $c(A) = a$ and $c(B) = b$. Therefore we have neither $aP^*b$ nor $bP^*a$.

Now let us prove that the Strict ucr only removes cycles of size 2. Two cases are possible to have no $P^*$ relations for two elements a and b:

- Either both a and b are available and never chosen. Therefore there is no revealed preference nor strictly unambiguous choice.
- Or, we must have some sets A and B such that $c(A) = a$ and $c(B) = b$, with $a, b \in A \cap B$. From A, we have that $aR^0b$ and from B, we have that $bR^0a$. So that we have a cycle of revealed preferences of size 2 between a and b, which will not be in $P^*$.

The previous reasoning proves that $P^*$ remove only cycles of size 2 of revealed preferences (something that is also a way to prove the equivalence with warp). Now as Strict ucr+ takes the Strict ucr and removes all cycles left, the proof is completed. ■

| Category | Subcategory | Percentage |
|---|---|---|
| Age of household head(s) | <35 | 26% |
| | $35 \leq$ age < 65 | 53% |
| | $\geq$ 65 | 21% |
| Household composition | 1-2 members | 49% |
| | 3-4 members | 38% |
| | >4 members | 14% |
| Number of household heads | 1 head | 23% |
| | 2 heads | 77% |
| Household income | < 20k | 13% |
| | 20k $\leq$ income < 45k | 57% |
| | $\geq$ 45k | 30% |
| Education of household head(s) | No College | 77% |
| | College | 23% |

Table 6.1: Summary of demographic characteristics, percentage are averaged, so that the sum might not be 100%.

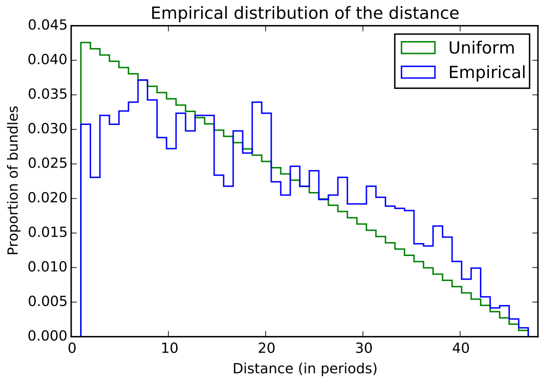



Figure 6.1: Probability density function of the distance between bundles in a cycle, compared to a uniformly random drawn distribution of them.

## 6.2 Appendix: Demographic characteristics of the sample

## 6.3 Appendix: Robustness Check, Stability of Preferences across time

To check the validity of the hypothesis of stability of preferences, we compute and plot the distribution of the distance in periods between two bundles, conditional on the fact that they are in a cycle. The probability density function obtained is in figure 6.1. It is significantly diﬀerent from a uniform distribution of errors across time, as told by the Kolmogorov-Smirnov (KS statistics of 0.09 and p-value of $10^{-11}$.). The distribution is slightly skewed towards longer distance. Households are less inconsistent between one period and the next one than between further away periods. The diﬀerence as shown in the figure does not seem to be of great concern.

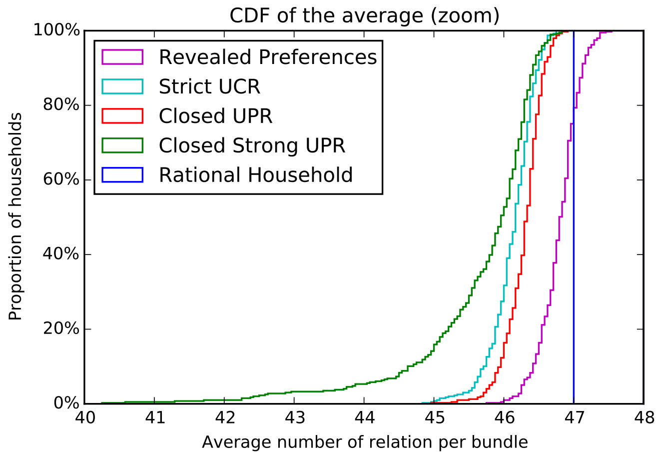



Figure 6.2: Average number of relation per bundle per household.

## 6.4 Appendix: Results while removing all cycles

One interpretation of the results of figure 6.2 is the following. Strict UCR is the relation we have talked about here. Strong UPR is the relation that consider all the revealed preferences relations and then remove all relations involved in a cycle. The transitive closure is shown here, but it does not change much. UPR is close to the minimal that can be achieved while being constrained to remove cycles and not relations only. It shows that it is much more costly that Strict UCR to remove all cycles, but that they are also some small gains in considering the removal of cycles directly from revealed preferences and not after Strict UCR.

## 6.5 Appendix: Computing the Selten's score

An average area and the Selten's score is computed for each household and each time period (that is, each observation). For each observation, a thousand random draws are performed. The random draws are used to assess a sort of "size" of the area on the hyperplane where a bundle would pass. The draws are performed in two steps:

1. A first draw to see whether the good has been chosen, using a empirical Bernoulli distribution. The distributions are good and household specific. The Bernoulli parameter is computed using the 48 time periods, when the good has been chosen, its value is 1, and when it does not, it is 0 (we assume independence of the choices at each time period). Then the values are averaged.

2. Once we now which goods have been virtually chosen, a uniform random draws on the budget shares of each good is performed, using the Dirichlet distribution, with parameters 1 for all good that are chosen. This distribution is the one that should be used to draw uniformly from the simplex, see

3. Quantities are then retrieved using the observed budget and prices for that period and that household. This allow for the test of GARP.

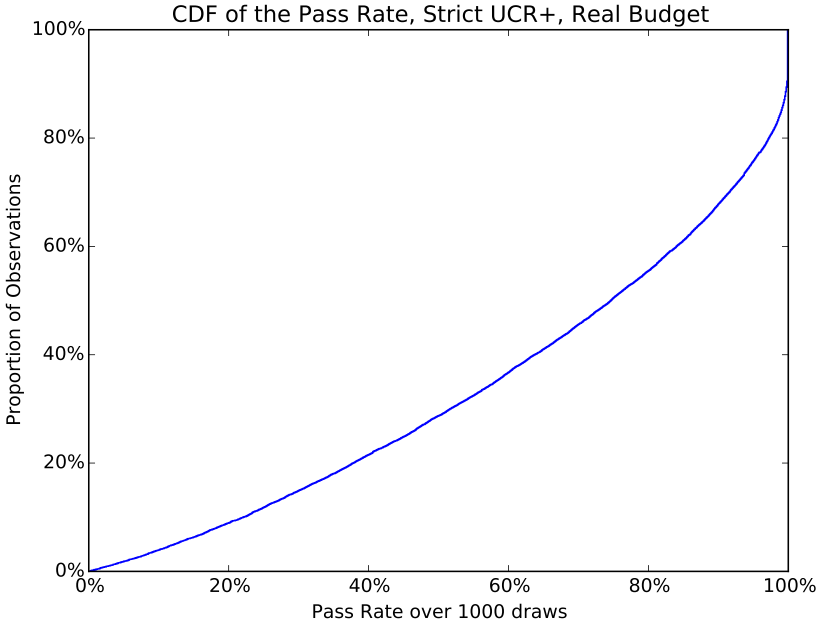



Figure 6.3: Disaggregated pass rate using real budget.

The reason why we use this indirect techniques and not directly uniform draws on the simplex for the budget shares is because a lot of goods are not chosen at each time period, thus generating corner solution. Using a uniform random draw with the real budgets in that case yields to stark results, with numerous areas being either 0 or 1. In average, for each bundle, 111 household have areas of pass of that are exactly 1, which correspond to 1/4 of the sample using pure uniform draws, whereas the same average is 32 with the two step method. The two step method is closer to the observation, because picking randomly an observed good, the probability that this good has not been chosen in that observation is 82.5%.

## 6.6 Appendix: Results with real budget

The results of the prediction using real budget are shown in figure 6.3. The main conclusion here is that there are much more observation with high pass rate, that is, prediction are not really tight. A random bundle on the real budget (the budget actually consumed in that period) is more likely to pass. This comes from the fact that budget lines are not crossing enough and are not restrictive.