Testable Implications of Models of Intertemporal Choice: Exponential Discounting and Its Generalizations [1]

Federico Echenique   Taisuke Imai   Kota Saito

California Institute of Technology

April 13, 2015

[1]We thank Kim Border and Chris Chambers for inspiration and advice. This paper subsumes the paper "Testable Implication of Exponential Discounting" (2014) circulated as Caltech Social Science Working Paper 1381. The authors wish to thank Jim Andreoni and Charlie Sprenger for their many detailed comments on a previous draft, and discussions that really benefited our paper. We are also grateful for the feedback provided by seminar audiences in the many different places where we have presented the paper. Imai is grateful for financial support from the Nakajima Foundation.

Abstract

We present the first revealed-preference characterizations of the most common models of intertemporal choice: the model of exponentially discounted concave utility, and some of its generalizations. Ours is the *first axiomatization* of these models taking *consumption data* as primitives. Our characterizations provide non-parametric revealed-preference tests. We apply our test to data from a recent experiment, and find that our axiomatization delivers new insights and perspectives on a dataset that had been analyzed by traditional parametric methods.

1. Introduction

Exponentially discounted utility is the standard model of intertemporal choice in economics. It is a ubiquitous model; used in all areas of economics. Our paper is the first revealed preference investigation of exponential discounting: We give a necessary and sufficient "revealed preference axiom" that a dataset must satisfy in order to be consistent with exponential discounting. The revealed preference axiom sheds light on the behavioral assumptions underlying the standard model of discounting. It also yields a non-parametric test of the theory, applicable in different empirical investigations of exponential discounting.

Consider an agent who chooses among intertemporal consumptions of a single good. One general theory is that the agent has a utility function $U(x_0, \dots, x_T)$ for the consumption of x_t on each date t. The Generalized Axiom of Revealed Preference (GARP) tells us whether the choices are consistent with some general utility function U.

The empirical content of general utility maximization is well understood, but it is too broad (and GARP is too weak) to capture exponential discounting. The *exponentially discounted utility* (EDU) model assumes a specific form of U, namely

$$U(x_0, \dots, x_T) = \sum_{t=0}^{T} \delta^t u(x_t).$$

In this paper, we focus on concave EDU, in which u is a concave function. Concavity of u is widely used to capture a motive for consumption smoothing over time. The empirical content of concave EDU maximization is different from that of general utility maximization, and not well understood in the literature.

The first and most important question addressed in our paper is: *What is the version of GARP that allows us to decide whether data are consistent with concave EDU?* The revealed preference axiom that characterizes concave EDU is obviously going to be stronger than GARP. Despite the ubiquity of EDU in economics, the literature on revealed preference has not (until now) provided an answer. Our main result is that a certain revealed preference axiom, termed the "Strong Axiom of Revealed Exponentially Discounted Utility" (SAR-EDU), describes the choice data that are consistent with concave EDU preferences.

SAR-EDU is a weak imposition on the data, in the sense that it constrains prices and quantities *in those situations in which unobservables do not matter.* The constraint on prices and quantities is simply that they be inversely related, or that "demand slopes down." Essentially, SAR-EDU requires one to consider situations in which unobservables "cancel out," and to check that prices and quantities are inversely related. This inverse relation is a basic implication of concave utility (that is, of the consumption smoothing motive).

In the paper, we study the empirical content of more general models of time discounting as well, including the quasi-hyperbolic discounting model (QHD; Phelps and Pollak, 1968; Laibson, 1997): $U(x_0, \ldots, x_T) = u(x_0) + \beta \sum_{t=1}^{T} \delta^t u(x_t)$, general time discounting (GTD): $U(x_0, \ldots, x_T) = \sum_{t=0}^{T} D(t) u(x_t)$, and time separable utility (TSU): $U(x_0, \ldots, x_T) = \sum_{t=0}^{T} u_t(x_t)$; where u and u_t are concave. In the following, we do not explicitly use the concave modifier when there is no risk of confusion. For example, we say EDU to mean concave EDU.

The contribution of this paper is to characterize the empirical content of EDU and its generalizations. We provide the first revealed-preference axioms (axioms like GARP but stronger) characterizing EDU, QHD, GTD, and TSU. Our axioms shed new insights into the behavioral assumptions behind each of these models, and also constitute non-parametric tests. There are, of course, other axiomatizations of these models but they start from different primitives. The well know axiomatization of EDU by Koopmans (1960), for example, starts from complete preferences over infinite consumption streams.

To illustrate the usefulness of our results for empirical work, we carry out an application to data from a recent experiment conducted by Andreoni and Sprenger (2012) (hereafter AS). AS propose the Convex Time Budget (CTB) method, in which subjects are asked to choose from an intertemporal budget set. [1] They find moderate support for the theory that agents are EDU maximizers.

The application of our methods to AS's data is, we believe, fruitful. We uncover features of individual subjects' behavior that are masked by traditional parametric econometric techniques. Our tests give a seemingly different conclusion from that obtained by AS. At first glance, we find scant evidence for EDU (or indeed QHD) whereas AS are cautiously supportive of EDU. Section 5 has more details, and reveals that the methodology of AS and our methodology are more concordant than what initially emerges.

It should be said that our methods rest on nonparametric revealed preference tests. As such, the tests are independent of functional form assumptions. The tests are also simple, and tightly connected to economic theory. The methodology used currently by experimentalists (such as AS) rests instead on parametrically estimating a given utility function by statistical methods. Our setup fits the experimental design of AS, and other CTB experiments, very well, but our results are also applicable more broadly, including to non-experimental field

[1] Several recent experimental studies use the CTB design, both in the laboratory and in the field setting, including Andreoni et al. (2013b), Ashton (2014), Augenblick et al. (forthcoming), Barcellos and Carvalho (2014), Brocas et al. (2015), Carvalho et al. (2013), Carvalho et al. (2014), Giné et al. (2013), Janssens et al. (2013), Kuhn et al. (2014), Liu et al. (2014), Lührmann et al. (2014), Sawada and Kuroishi (2015), and Shaw et al. (2014). Our methods are largely applicable to data from these experiments.

data.

Related literature. There are different behavioral axiomatizations of EDU in the literature, starting with Koopmans (1960), and followed by Fishburn and Rubinstein (1982) and Fishburn and Edwards (1997). All of them take preferences as primitive, or in some cases they take utility over consumption streams as the primitive. The idea is that the relevant behavior consists of all pairwise comparisons of consumption streams. From an empirical perspective, this assumes a infinite "dataset" of pairwise comparisons. Indeed the stationarity axiom introduced by Koopmans (1960), and used by many other authors, requires infinite time. Our axiomatization of EDU is the first in an environment where agents choose from budget sets.

Other axiomatizations of EDU impose stationarity in different environments. In Fishburn and Rubinstein (1982) preferences are defined on one-time consumptions in continuous time. In Fishburn and Edwards (1997), preferences are defined on infinite consumption streams that differ in at most finitely many periods. The recent work of Dziewulski (2014) gives a characterization for binary comparisons of one-time consumptions, a similar setup to Fishburn and Rubinstein, but assuming finitely many data.

In continuous time setup, Weibull (1985) gives a general characterization of EDU, also taking preferences as primitives. He characterizes the general time discounting models and the monotone time discounting models as well. A more recent paper by Kopylov (2010) also provides a simple axiomatization of EDU model in a continuous time setup.

The QHD model was first proposed by Phelps and Pollak (1968), who did not propose an axiomatization. There are several more recent studies that present a behavioral characterization of QHD, but all take preferences and infinite time horizons as their primitives, and therefore differ from our results. See Hayashi (2003), Montiel Olea and Strzalecki (2014), and Galperti and Strulovici (2014) for axiomatizations.

Time separable utility (TSU) model is the most general model we axiomatize. In our application of our test to AS's data, however, we found that significant number of subjects are not TSU rational. This would suggest the importance of non time separable model. Gilboa (1989) has provided an elegant axiomatization of a non time separable utility model. In the paper, by using Anscombe and Aumann's framework and studying preferences over finite sequences of lotteries, Gilboa (1989) axiomatizes a utility function that can capture a preference for (or an aversion to) variation of utility levels across periods.

In terms of data from (field) consumption surveys, Browning (1989) provides a revealed-preference axiom for EDU with no discounting (i.e., $\delta = 1$). Other papers on survey data do not provide an axiomatic characterization; they, instead, obtain Afriat inequalities for several models. Crawford (2010) investigates intertemporal consumption and discusses a particular violation of TSU, namely habit formation. Crawford presents Afriat inequalities for the model of habit formation, and uses Spanish consumption data to carry out the test (see also Crawford and Polisson, 2014). Adams et al. (2014) work with the Spanish dataset and test EDU within a model of collective decision making at the household level.

It is important to emphasize that the papers on survey data allow for the existence of many goods in each period; but they do not allow for more than one (intertemporal) purchase for each agent. This assumption makes sense because in consumption surveys one typically has a single observation per household. We have instead assumed that there is only one good (money) in each period; but we allow for more than one intertemporal purchase per agent. Allowing for multiple purchases is crucial in order to apply our tests to experimental data. This is because in experiments, a subject is usually required to make many decisions and one choice is chosen randomly for the payment to the subject.

2. Exponentially Discounted Utility

Notational conventions. For vectors $x, y \in \mathbf{R}^n$, $x \leq y$ means that $x_i \leq y_i$ for all $i = 1, \ldots, n$; $x < y$ means that $x \leq y$ and $x \neq y$; and $x \ll y$ means that $x_i < y_i$ for all $i = 1, \ldots, n$. The set of all $x \in \mathbf{R}^n$ with $0 \leq x$ is denoted by $\mathbf{R}^n_+$ and the set of all $x \in \mathbf{R}^n$ with $0 \ll x$ is denoted by $\mathbf{R}^n_{++}$.

Let T be a strictly positive integer; T will be the (finite) duration of time, or *time horizon*. We abuse notation and use T to denote the set $\{0, 1, \ldots, T\}$. A sequence $(x_0, \ldots, x_T) = (x_t)_{t \in T} \in \mathbf{R}^T_+$ will be called a *consumption stream*. There is a single good in each period; the good can be thought of as money. Note that the cardinality of the set T is $T + 1$, but this never leads to confusion.

Remark 1. We can assume a more general time setup: $\{0, \tau_1, \ldots, \tau_T\}$, where $\tau_i < \tau_{i+1}$ for all $i < T - 1$. Even with this general time setup, our results hold without changes. The only requirement on the set of time periods is that it contains 0. Such flexibility in how one specifies time is necessary in the application of our results to experimental data of Andreoni and Sprenger (2012). See Section 5.1 for detail.

The model. The objects of choice in our model are consumption streams. We assume that an agent has a budget $I > 0$, faces prices $p \in \mathbf{R}^T_{++}$, and chooses an affordable consumption stream $(x_t)_{t \in T} \in \mathbf{R}^T_+$. Prices can be thought of as interest rates.

An exponentially discounted utility (EDU) is specified by a discount factor $\delta \in (0,1]$ and a utility function over money $u : \mathbf{R}_+ \to \mathbf{R}$. An EDU maximizing agent solves the problem

$$\max_{x \in B(p,I)} \sum_{t \in T} \delta^t u(x_t) \tag{1}$$

when faced with prices $p \in \mathbf{R}^T_{++}$ and budget $I > 0$. The set $B(p, I) = \{y \in \mathbf{R}^T_+ : p \cdot y \leq I\}$ is the *budget set* defined by p and I.

The meaning of EDU as an assumption about an agent is that the agent's observed behavior is *as if* it were generated by the maximization of an EDU. To formalize this idea, we need to state what can be observed.

Definition 1. *A* dataset *is a finite collection of pairs* $(x, p) \in \mathbf{R}^T_+ \times \mathbf{R}^T_{++}$.

A dataset is our notion of observable behavior. The interpretation of a dataset $(x^k, p^k)^K_{k=1}$ is that it describes K observations of a consumption stream $x^k = (x^k_t)_{t \in T}$ at some given vector of prices $p^k = (p^k_t)_{t \in T}$, and budget $p^k \cdot x^k = \sum_{t \in T} p^k_t x^k_t$. We sometimes use K to denote the set $\{1, \ldots, K\}$.

Let us clarify the meaning of a dataset by considering two examples. If we have field consumption data, collected through a consumption survey, then K is 1. There is one dataset for each agent, or household. This is the setup of Browning (1989), for example. On the other hand, if, in an experiment, one subject is asked to make a choice from 45 different budget sets, as in Andreoni and Sprenger (2012), then K is 45. The experimenter would typically implement the choice from one budget set selected at random. It is important to note that our model allows, but does not require, that $K > 1$. Even if $K = 1$, our axiom for EDU is not satisfied trivially, and has testable implications.

Definition 2. *A dataset* $(x^k, p^k)^K_{k=1}$ *is* exponential discounted utility rational *(EDU rational) if there is* $\delta \in (0, 1]$ *and a concave and strictly increasing function* $u : \mathbf{R}_+ \to \mathbf{R}$ *such that, for all* k*,*

$$y \in B(p^k, p^k \cdot x^k) \Longrightarrow \sum_{t \in T} \delta^t u(y_t) \leq \sum_{t \in T} \delta^t u(x^k_t).$$

As mentioned in the introduction, we restrict attention to concave utility. Our results will be silent about the non-concave case. So we are focusing on agents who seek to smooth out their consumption over time.

3. A characterization of EDU rational data

EDU rational data will be characterized by a single "revealed preference axiom." We shall introduce the axiom by deriving the implications of EDU in specific instances. Here we assume, for ease of exposition, that u is differentiable, but our results do not depend on differentiability, and the statement of the theorem will not require the differentiability of u.

The first-order condition for maximization of EDU is: for each $k \in K$ and $t \in T$,

$$\delta^t u'(x_t^k) = \lambda^k p_t^k. \tag{2}$$

The first-order conditions involve three unobservables: discount factor δ, marginal utilities $u'(x_t^k)$ and Lagrange multipliers λ^k. Quantities x_t^k and prices p_t^k are observable. Our approach proceeds by finding that certain implications of the model for the observables x^k and p^k must hold, regardless of the values of the unobservables.

We derive the axiom by considering increasingly general cases. First we consider the case of no discounting and one observation ($\delta = 1$ and $K = 1$). Then, we study the case of no discounting ($\delta = 1$ and $K \geq 1$). Finally, in Section 3.3 we discuss the general case (δ is unknown and $K \geq 1$) and present the axiom for EDU, SAR-EDU.

3.1. No discounting and one observation: $\delta = 1$ and $K = 1$

Suppose that $\delta = 1$ and $K = 1$. That is, we seek to impose EDU rationality in the special case when δ is known, equals 1, and our dataset has a single observation. Under these assumptions (omitting the k superindex, as $K = 1$) the first-order condition (2) becomes $u'(x_t) = \lambda p_t$ for each $t \in T$. For each pair $t, t' \in T$, we obtain

$$\frac{u'(x_t)}{u'(x_{t'})} = \frac{p_t}{p_{t'}}.$$

By concavity of u, for each pair $t, t' \in T$, we have

$$x_t > x_{t'} \Longrightarrow \frac{p_t}{p_{t'}} \leq 1. \tag{3}$$

Thus we obtain a simple implication of EDU rationality in this special case: (3) means that *demand must slope down.* This "downward sloping demand axiom" coincides with the axiom obtained by Browning (1989) for the $\delta = K = 1$ case.[2]

Property (3) can be written in a different way. It is more complicated than (3), and redundant for now, but will prove useful in the sequel:

Definition 3. *A sequence of pairs* $(x_{t_i}^{k_i}, x_{t'_i}^{k'_i})_{i=1}^n$ *has the* downward sloping demand property *if*

$$x_{t_i}^{k_i} > x_{t'_i}^{k'_i} \text{ for all } i \text{ implies that } \prod_{i=1}^n \frac{p_{t_i}^{k_i}}{p_{t'_i}^{k'_i}} \leq 1.$$

[2] Browning is interested in the case of $K = 1$ because he uses survey consumption data.

The downward-sloping demand property is not only a necessary condition, but also a sufficient condition for EDU rationality in the case of $\delta = 1$ and $K = 1$.

3.2. No discounting: $\delta = 1$

We now take one step towards our general result. Continue to assume that $\delta = 1$, but now allow that $K \geq 1$. The decision maker does not discount future consumptions, but the dataset may contain multiple observations. The first-order condition (2) becomes $u'(x_t^k) = \lambda^k p_t^k$ for each $t \in T$, and each k.

If we try to proceed as in the previous section, we might consider pairs of observations with $x_t^k > x_{t'}^{k'}$:

$$\frac{u'(x_t^k)}{u'(x_{t'}^{k'})} = \frac{\lambda^k}{\lambda^{k'}} \frac{p_t^k}{p_{t'}^{k'}}.$$

By the concavity of u, we know that

$$x_t^k > x_{t'}^{k'} \implies \frac{\lambda^k}{\lambda^{k'}} \frac{p_t^k}{p_{t'}^{k'}} \leq 1.$$

However, the ratio $\lambda^k/\lambda^{k'}$ does not allow us to conclude anything about the ratio of prices. We would like to conclude, along the lines of downward sloping demand, that $p_t^k/p_{t'}^{k'} \leq 1$. But the presence of $\lambda^k/\lambda^{k'}$ does not allow us to do that. Of course if we consider $x_t^k > x_{t'}^{k'}$ for the same observation: $k = k'$ then the conclusion of downward sloping demand continues to hold. When $K > 1$, downward sloping demand is still a restriction within each observation k.

This suggest that we can obtain an implication of EDU (with $\delta = 1$) across observations as well. Consider a collection of pairs $(x_t^k, x_{t'}^{k'})$, chosen such that the λ variables will cancel out. For example consider:

$$\frac{u'(x_{t_1}^k)}{u'(x_{t_2}^{k'})} \frac{u'(x_{t_3}^{k'})}{u'(x_{t_4}^k)} = \frac{\lambda^k}{\lambda^{k'}} \frac{\lambda^{k'}}{\lambda^k} \frac{p_{t_1}^k}{p_{t_2}^{k'}} \frac{p_{t_3}^{k'}}{p_{t_4}^k}.$$

Then the λ variables cancel out and we obtain that:

$$x_{t_1}^k > x_{t_2}^{k'} \text{ and } x_{t_3}^{k'} > x_{t_4}^k \implies \frac{p_{t_1}^k}{p_{t_2}^{k'}} \frac{p_{t_3}^{k'}}{p_{t_4}^k} \leq 1;$$

that is, downward sloping demand.

The idea of canceling out the unknown λ^ks suggests the following definition.

Definition 4. *A sequence of pairs* $(x_{t_i}^{k_i}, x_{t'_i}^{k'_i})_{i=1}^n$ *is* balanced *if each* k *appears as* k_i *(on the left of the pair) the same number of times it appears as* k'_i *(on the right).*

When $K = 1$ we know that a sequence must have the downward sloping demand property. Now with $K \geq 1$ this is only true of *balanced* sequences: *any balanced sequence has the downward sloping demand property.* This property is not only necessary, but also a sufficient condition for EDU rationality in the case when δ is known and $\delta = 1$.

3.3. General K and δ

We now turn to the case when K can be arbitrary and δ is unknown. Before, when $K = \delta = 1$, then λ was constant and δ fixed. EDU rationality is characterized by downward sloping demand. When $K \geq 1$ we saw that we needed to impose downward sloping demand for balanced sequences. When δ is unknown we need to further restrict the sequences that are required to satisfy downward sloping demand. In fact, the relevant axiom turns out to be:

Strong Axiom of Revealed Exponentially Discounted Utility (SAR-EDU): *For any balanced sequence of pairs* $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^n_{i=1}$, *if* $\sum_{i=1}^n t_i \geq \sum_{i=1}^n t'_i$, *then the sequence has the downward sloping demand property.*

As in Sections 3.1 and 3.2, the key idea is to control the effects of the unknowns u, δ and λ, by focusing on particular configurations of the data. It is easy to see how such restrictions are necessary. For example, consider two pairs

$$(x^{k_1}_{t_1}, x^{k_2}_{t_2}) \text{ and } (x^{k_2}_{t_3}, x^{k_1}_{t_4})$$

such that

$$t_1 + t_3 \geq t_2 + t_4.$$

By manipulating first-order conditions we obtain that:

$$\frac{u'(x^{k_1}_{t_1})}{u'(x^{k_2}_{t_2})} \cdot \frac{u'(x^{k_2}_{t_3})}{u'(x^{k_1}_{t_4})} = \left(\frac{\delta^{t_2}}{\delta^{t_1}} \frac{\lambda^{k_1} p^{k_1}_{t_1}}{\lambda^{k_2} p^{k_2}_{t_2}} \right) \cdot \left(\frac{\delta^{t_4}}{\delta^{t_3}} \frac{\lambda^{k_2} p^{k_2}_{t_3}}{\lambda^{k_1} p^{k_1}_{t_4}} \right) = \delta^{(t_2+t_4)-(t_1+t_3)} \frac{p^{k_1}_{t_1}}{p^{k_2}_{t_2}} \frac{p^{k_2}_{t_3}}{p^{k_1}_{t_4}}.$$

Notice that the pairs $(x^{k_1}_{t_1}, x^{k_2}_{t_2})$ and $(x^{k_2}_{t_3}, x^{k_1}_{t_4})$ constitute a balanced sequence of pairs, so that the Lagrange multipliers cancel out as in 3.2. Furthermore, the discount factors unambiguously increase the value on the left hand side, $\delta^{(t_2+t_4)-(t_1+t_3)} \geq 1$ for any $\delta \in (0, 1]$.

Now the concavity of u implies that when $x^{k_1}_{t_1} > x^{k_2}_{t_2}$ and $x^{k_2}_{t_3} > x^{k_1}_{t_4}$ then the product $\delta^{(t_2+t_4)-(t_1+t_3)}(p^{k_1}_{t_1}/p^{k_2}_{t_2})(p^{k_2}_{t_3}/p^{k_1}_{t_4})$ cannot exceed 1. Since $\delta^{(t_2+t_4)-(t_1+t_3)} \geq 1$ for any $\delta \in (0, 1]$, then $(p^{k_1}_{t_1}/p^{k_2}_{t_2})(p^{k_2}_{t_3}/p^{k_1}_{t_4})$ cannot exceed 1. Thus, we obtain an implication of EDU for prices, an observable entity. *No matter what the values of the unobservable δ and u, we find that the ratio of prices cannot be more than* 1.

The argument just made extends to arbitrary balanced sequences, and essentially gives the proof of necessity of SAR-EDU.[3] The argument simply amounts to verifying a rather basic consequence of EDU: the consequence of EDU for those situations in which unobservables either do not matter or have a known effect (the effect either resulting from u' being decreasing or from $\delta \in (0,1]$). What is surprising is that such a basic consequence of the theory is sufficient as well as necessary.

Theorem 1. *A dataset is EDU rational if and only if it satisfies SAR-EDU.*

The proof is in Section 6. The proof that SAR-EDU is necessary is, as we have remarked, simple. The proof of sufficiency is more complicated, and follows ideas introduced in Echenique and Saito (forthcoming).

Remark 2. It is not obvious from the syntax of SAR-EDU that one can verify whether a particular dataset satisfies SAR-EDU in finitely many steps. We can show that, not only is SAR-EDU decidable in finitely many steps, but there is in fact an efficient algorithm that decides whether a dataset satisfies SAR-EDU. The proof is very similar to Proposition 2 in Echenique and Saito (forthcoming). So we omit the proof. SAR-EDU is on the same computational standing as GARP or the strong axiom of revealed preference. Another way to test SAR-EDU is based on linearized "Afriat inequalities," see Lemma 1 of Section 6.3. In fact, this is how we proceed in Section 5; see in particular the discussion at the end of that section.

4. More general models

The ideas behind Theorem 1 can be used to analyze other models of intertemporal choice, including quasi-hyperbolic discounting (QHD), and more general models.

4.1. Quasi-Hyperbolic Discounted Utility

First we investigate QHD. The objective is the same as for EDU: we want to know when a dataset $(x^k, p^k)_{k=1}^K$ is consistent with QHD utility maximization, but the interpretation of a dataset is now more complicated. In the case of QHD, we assume that each x^k is a consumption stream that the agent *commits* to at date 0. The reason is that a QHD agent may be dynamically inconsistent, and revise their planned consumption. The commitment assumption happens to perfectly fit the application in Section 5 to the CTB experiment in

[3] We have assumed differentiability of u in our informal derivation, but since u is concave, we can easily generalize the argument.

Andreoni and Sprenger (2012). The commitment assumption will, however, be violated by field data taken from consumption surveys. It is important to note that the assumption of commitment is not necessary to test the EDU model, which is dynamically consistent.

Definition 5. *A dataset $(x^k, p^k)_{k=1}^K$ is* quasi-hyperbolic discounted utility rational *(QHD rational) if there is $\delta \in (0,1]$, index for time-bias $\beta > 0$, and a concave and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$ such that, for all k,*

$$y \in B(p^k, p^k \cdot x^k) \Longrightarrow \sum_{t\in T} D(t)u(y_t) \le \sum_{t\in T} D(t)u(x_t^k),$$

where $D(t) = 1$ if $t = 0$ and $D(t) = \beta\delta^t$ if $t > 0$. More specifically, if $\beta \le 1$ in the above definition then the dataset $(x^k, p^k)_{k=1}^K$ is present biased quasi-hyperbolic discounted utility rational *(PQHD-rational).*

Strong Axiom of Revealed Quasi-Hyperbolic Discounted Utility (SAR-QHD): *For any balanced sequence of pairs $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$, if*

1. $\sum_{i=1}^n t_i \ge \sum_{i=1}^n t_i'$ *and*
2. $\#\{i : t_i > 0\} = \#\{i : t_i' > 0\}$,

then the sequence has the downward sloping demand property.

The condition that $\sum_{i=1}^n t_i \ge \sum_{i=1}^n t_i'$ plays the same role as it did in SAR-EDU, to control the effect of δ. In addition, we must now have $\#\{i : t_i > 0\} = \#\{i : t_i' > 0\}$ so as to cancel out β. If we instead focus on PQHD, then we know that $\beta \le 1$ so the weaker requirement $\#\{i : t_i > 0\} \ge \#\{i : t_i' > 0\}$ controls the effect of β.[4] Formally, the axiom to characterize PQHD is as follows:

Strong Axiom of Revealed Quasi-Hyperbolic Present-Biased Utility (SAR-PQHD): *For any balanced sequence of pairs $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$, if*

1. $\sum_{i=1}^n t_i \ge \sum_{i=1}^n t_i'$ *and*
2. $\#\{i : t_i > 0\} \ge \#\{i : t_i' > 0\}$,

then the sequence has the downward sloping demand property.

[4] It is easy to axiomatize *future biased quasi-hyperbolic discounted utility* (FQHD), which is a special model of QHD with $\beta \ge 1$. For FQHD, in turn, we need $\#\{i : t_i > 0\} \le \#\{i : t_i' > 0\}$.

To show the necessity of SAR-QHD, we proceed as in Section 3.3. The first order conditions for maximization of a QHD utility are:

$$D(t)u'(x_t^k) = \lambda^k p_t^k. \tag{4}$$

For example consider a balanced sequence of pairs $(x_{t_1}^{k_1}, x_{t_2}^{k_2})$, $(x_{t_3}^{k_2}, x_{t_4}^{k_1})$ with the property that $t_1 + t_3 \geq t_2 + t_4$, and $\#\{i \in \{1,3\} : t_i > 0\} = \#\{i \in \{2,4\} : t_i > 0\}$; where $\#\{i \in \{1,3\} : t_i > 0\}$ and $\#\{i \in \{2,4\} : t_i > 0\}$ are the numbers of non-time-zero consumption in $\{x_{t_1}^{k_1}, x_{t_3}^{k_3}\}$ and $\{x_{t_2}^{k_2}, x_{t_4}^{k_4}\}$, respectively. By manipulating the first-order conditions we obtain that:

$$\begin{aligned}
\frac{u'(x_{t_1}^{k_1})}{u'(x_{t_2}^{k_2})} \cdot \frac{u'(x_{t_3}^{k_2})}{u'(x_{t_4}^{k_1})} &= \left(\frac{\beta^{1_{\{t_2>0\}}}}{\beta^{1_{\{t_1>0\}}}} \frac{\delta^{t_2}}{\delta^{t_1}} \frac{\lambda^{k_1} p_{t_1}^{k_1}}{\lambda^{k_2} p_{t_2}^{k_2}} \right) \cdot \left(\frac{\beta^{1_{\{t_4>0\}}}}{\beta^{1_{\{t_3>0\}}}} \frac{\delta^{t_4}}{\delta^{t_3}} \frac{\lambda^{k_2} p_{t_3}^{k_2}}{\lambda^{k_1} p_{t_4}^{k_1}} \right) \\
&= \beta^{\#\{i \in \{2,4\}: t_i > 0\} - \#\{i \in \{1,3\}: t_i > 0\}} \delta^{(t_2+t_4)-(t_1+t_3)} \frac{p_{t_1}^{k_1}}{p_{t_2}^{k_2}} \frac{p_{t_3}^{k_2}}{p_{t_4}^{k_1}} \\
&= \delta^{(t_2+t_4)-(t_1+t_3)} \frac{p_{t_1}^{k_1}}{p_{t_2}^{k_2}} \frac{p_{t_3}^{k_2}}{p_{t_4}^{k_1}}.
\end{aligned}$$

The balancedness of the sequence of pairs $(x_{t_1}^{k_1}, x_{t_2}^{k_2})$ and $(x_{t_3}^{k_2}, x_{t_4}^{k_1})$ implies that Lagrange multipliers cancel out. The assumption of $\#\{i \in \{1,3\} : t_i > 0\} = \#\{i \in \{2,4\} : t_i > 0\}$ implies that β cancels out. As in SAR-EDU, the discount factor unambiguously increases the value of the right hand side.

Finally, concavity of u implies that, when $x_{t_1}^{k_1} > x_{t_2}^{k_2}$ and $x_{t_3}^{k_2} > x_{s_4}^{k_1}$, we have that $(p_{t_1}^{k_1}/p_{t_2}^{k_2})(p_{t_3}^{k_2}/p_{t_4}^{k_1})$ cannot exceed 1. That is, the downward sloping demand property.

The next theorem summarizes our results on QHD.

Theorem 2. *A dataset is QHD-rational if and only if it satisfies SAR-QHD. Moreover, the dataset is PQHD-rational if and only if it satisfies SAR-PQHD.*

The proof of Theorem 2 is in Section 7.

One consequence of Theorems 1 and 2 is that, under certain circumstances, EDU and PQHD are *observationally equivalent*. These circumstances are very relevant for the discussion in Section 5 of AS's experiment: Our next result, Proposition 1, shows that if an agent does not consume at the soonest date (i.e., $x_0^k = 0$ for all $k \in K$), then EDU and PQHD are observationally equivalent. In AS's experiment, 82.8% of the subjects (i.e., 24 out of 29 subjects) who satisfy SAR-EDU do not consume at the soonest date. This explains why, in AS's data, QHD has very limited scope beyond what can be explained by EDU.

Proposition 1. *Suppose that a dataset $(x^k, p^k)_{k=1}^K$ satisfies that $x_0^k = 0$ for all $k \in K$. Then $(x^k, p^k)_{k=1}^K$ is PQHD rational if and only if it is EDU rational.*

Proof. Of course, if the data is EDU rational then it is PQHD rational. Let us prove the converse. Choose a sequence $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^n_{i=1}$ such that (1) $x^{k_i}_{t_i} > x^{k'_i}_{t'_i}$ for all $i \in \{1, \ldots, n\}$, (2) $\sum_{i=1}^n t_i \geq \sum_{i=1}^n t'_i$, and (3) each k appears as k_i the same number of times as k'_i.

By (1), $x^{k_i}_{t_i} > 0$ for all $i \in \{1, \ldots, n\}$. Since $x^k_0 = 0$ for all $k \in K$, we obtain $t_i > 0$ for all $i \in \{1, \ldots, n\}$. Therefore, $\#\{i \in \{1, \ldots, n\} : t_i > 0\} = \#\{i \in \{1, \ldots, n\}\} \geq \#\{i \in \{1, \ldots, n\} : t'_i > 0\}$. Therefore, the sequence satisfies all of the conditions in the strong axiom of present biased QHD. Since the dataset is PQHD rational, Theorem 2 shows that

$$\prod_{i=1}^{n} \frac{p^{k_i}_{t_i}}{p^{k_i}_{t'_i}} \leq 1. \tag{5}$$

Therefore, Conditions (1), (2), and (3) imply (5), which is SAR-EDU. Hence, by Theorem 1, the dataset must be EDU rational. □

4.2. More General Models of Time Discounting

Building on the ideas in the previous two theorems, we can characterize more general models of intertemporal choice. These models end up being useful in Section 5.2 when we classify subjects in AS's experiment.

Of course the most general model is utility maximization, without constraints on the form of the utility.

$$\max \; U(x_0, \ldots, x_T) \;\; \text{s.t.} \;\; p \cdot x \leq I. \tag{6}$$

The relevant revealed preference axiom is GARP. In the following, we provide three special cases of (6), which are obtained by restricting U. We list the three models in order of generality. Let $\mathcal{C}$ be the set of all continuous, concave, and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$.

1. *Time-separable utility* (TSU): The class TSU of all U that can be written as

$$U(x_0, \ldots, x_T) = \sum_{t \in T} u_t(x_t),$$

where $u_t \in \mathcal{C}$ for all $t \in T$.

2. *General time discounting* (GTD): The class GTD of all U that can be written as

$$U(x_0, \ldots, x_T) = \sum_{t \in T} D(t) u(x_t),$$

for some $u \in \mathcal{C}$, and a function $D : T \to \mathbf{R}_{++}$.

3. *Monotone time discounting* (MTD): The class MTD of all U that can be written as

$$U(x_0, \ldots, x_T) = \sum_{t \in T} D(t) u(x_t),$$

for some $u \in \mathcal{C}$, and a function $D : T \to \mathbf{R}_{++}$ that is monotonically decreasing.

In the following definition, the set M of utility functions can be any of the classes defined above (i.e., TSU, GTD, MTD)

Definition 6. *For $M \in \{TSU,\ GTD,\ MTD\}$, a dataset $(x^k, p^k)_{k=1}^K$ is* M rational *if there is a utility function U in the class M of utilities such that for all k,*

$$p^k \cdot y \leq p^k \cdot x^k \Longrightarrow U(y) \leq U(x^k).$$

It is easy to derive each axiom from first order conditions as we did for EDU and QHD. The idea is to choose a sequence of pairs of observations so that we can cancel out the Lagrange multipliers and control or cancel out the effects of other unobservables. We omit the derivations.

Strong Axiom of Revealed Time Separable Utility (SAR-TSU): *For any balanced sequence of pairs $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$, if each $t_i = t_i'$ for all i, then the sequence has the downward sloping demand property.*

Strong Axiom of Revealed General Time Discounted Utility (SAR-GTD): *For any balanced sequence of pairs $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$, if each t appears as t_i (on the left of the pair) the same number of times it appears as t_i' (on the right), then the sequence has the downward sloping demand property.*

Strong Axiom of Revealed Monotone Time Discounted Utility (SAR-MTD): *For any balanced sequence of pairs $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$, if there is a permutation π of $\{1, 2, \ldots, n\}$ such that $t_i \geq t'_{\pi(i)}$, then the sequence has the downward sloping demand property.*

Each of these axioms imposes the downward-sloping demand property of a balanced sequence under increasingly demanding conditions. For example, SAR-TSU imposes the downward sloping demand property of a subset of the sequences that are constrained by SAR-MTD; and SAR-MTD in turn constrains fewer sequences than SAR-EDU. How demanding an axiom is, in terms of imposing the downward-sloping property, mirrors how demanding the theory is: EDU is a special case of MTD, which is a special case of TSU.

Theorem 3. *Let $M \in \{TSU,\ GTD,\ MTD\}$. A dataset is M-rational if and only if it satisfies SAR-M.*

The proof of Theorem 3 follows similar ideas to those used in the proofs of the other two results, and is relegated to the Online Appendix.

5. Empirical Application

5.1. Description of the Data

AS introduce an experimental method called the Convex Time Budget (CTB) design. In contrast with the "multiple price list method" (e.g., Andersen et al., 2008), subjects in AS are asked to allocate 100 experimental tokens between "sooner" (time τ) and "later" (time $\tau + d$) accounts. Tokens allocated to each account have a value of a_τ and $a_{\tau+d}$, converting experimental currency unit into real monetary value for final payments. The gross interest rate over d days is thus given by $a_{\tau+d}/a_\tau$. There are three possible sooner dates $\tau \in \{0, 7, 35\}$; three possible delays $d \in \{35, 70, 98\}$ (the unit of period is one day.); and five different pairs of conversion rates $(a_\tau, a_{\tau+d})$. Each subject thus complete 45 decisions.[5]

Each subject's decision in a trial is characterized by a tuple $(\tau, d, a_\tau, a_{\tau+d}, c_\tau)$: the first four elements $(\tau, d, a_\tau, a_{\tau+d})$ characterize the budget set she faces in this trial; and c_τ is the number of tokens she decides to allocate to the sooner payment.

In the experiment, subjects make a two-period choice. They choose $(x_\tau, x_{\tau+d})$ subject to $p_\tau x_\tau + x_{\tau+d} = I$. We need to formulate the problem as choosing $(x_0, \dots x_T)$ subject to $\sum_{t \in T} p_t x_t = I$. We set prices to be $p_\tau = a_{\tau+d}/a_\tau$ and $p_{\tau+d} = 1$ (a normalization); and we define consumptions (monetary amounts) $x_\tau = c_\tau \cdot a_\tau$ and $x_{\tau+d} = (100 - c_\tau) \cdot a_{\tau+d}$.

We shall implicitly set the prices of periods that are not offered to be very high, so that agents choose zero consumption in those periods. For example, when subjects face a convex budget with $(\tau, d) = (35, 70)$, we treat prices p_t for $t \neq 35, 105$ as high. In any of the rationalizations we consider, marginal utilities at zero are finite. So by setting such prices high enough, the choices in such time periods do not affect whether a dataset is rationalizable. In this way, for each of 97 subjects, we obtained a dataset with $K = 45$ and $T = \{t : t = \tau \text{ or } t = \tau + d \text{ for some } \tau \in \{0, 7, 35\} \text{ and } d \in \{35, 70, 98\}\}$.

Three features of the AS design make their experiment ideal for our exercise. First and most importantly, the experimental setup is precisely the situation our model tries to capture: subjects choose an intertemporal consumption from a budget set. As we briefly mention above, most previous experimental studies on intertemporal decision utilize an environment with discrete (in many cases, binary) choice sets. Strictly speaking, budgets in AS experiment are discrete as well, but we understand them to be a reasonable approximation to continuous

[5] See Figure B.1 in the Online Appendix for an illustration. For each pair of starting date and delay length (τ, d), the 5 budgets are nested. Looking at all 45 budget sets, except for 8 cases in which $(a_\tau, a_{\tau+d}) = (0.2, 0.25)$, $a_{\tau+d}$ is fixed at 0.2 and a_τ ranges between 0.1 and 0.2. Participants' choices therefore always satisfy GARP by design.

choice (tokens are worth \$0.1 to \$0.25).

Secondly, the AS experiment has subjects *committing* to a consumption stream. Recall that to test for QHD and more general models (although not for EDU) we need to assume that agents commit to a consumption stream. In the AS design, the commitment assumption is satisfied.

Thirdly, AS put significant effort into equalizing the transaction costs of sooner and later payments, and minimizing the unwanted effects of uncertainty regarding future payments.

Before discussing our results, we summarize AS's main findings. AS estimate the per-period discount factor, present bias, and utility curvature assuming a QHD model with CRRA utility over money:

$$U(x_0, \ldots, x_T) = \frac{1}{\alpha} x_0^\alpha + \beta \sum_{t \in T \setminus \{0\}} \delta^t \frac{1}{\alpha} x_t^\alpha. \tag{7}$$

Their estimation uses pooled data from all subjects, fitting a common specification (7). AS find no evidence of present bias ($\widehat{\beta} = 1.007$, $SE = 0.006$; the hypothesis of no present bias, $\beta = 1$, is not rejected; $F_{1,96} = 1.51$, $p = 0.22$).[6] AS also estimate (7) at the level of individual subjects and find that the estimated $\widehat{\beta}$'s are narrowly distributed around 1, with the median estimate being 1.0011.

5.2. Results

We test our axioms for each individual subject in AS's experiment. Note that we do not pool the choice data of different subjects. The tests are based on the linearized Afriat inequalities presented in Lemma 1. The models we examine are EDU, QHD, MTD, GTD, and TSU. In the sequel, we shall label a subject as "M rational" if her dataset passes the revealed preference test for model M and "M non-rational" otherwise. The models can be ordered by the tightness of the associated axioms. Essentially, we have that EDU $\subset$ PQHD $\subset$ MTD $\subset$ GTD $\subset$ TSU, and that EDU $\subset$ QHD $\subset$ GTD $\subset$ TSU, as QHD is not comparable to MTD (QHD allows $\beta > 1$). For this reason, when we find that a subject is EDU rational, she is of course also M rational for all other models M $\in \{\text{PQHD}, \text{QHD}, \text{MTD}, \text{GTD}, \text{TSU}\}$. We sometimes label a subject as "strictly M rational" for the most restrictive model M such

[6] AS estimate several model specifications (e.g., assuming CARA instead of CRRA, or incorporating additional parameters to capture background consumptions), and they also use different estimation methods (e.g., two-limit Tobit model to handle corner choices). In our comparison, we use their results from a nonlinear least squares estimation of quasi-hyperbolic discounting and CRRA utility function without background consumption.

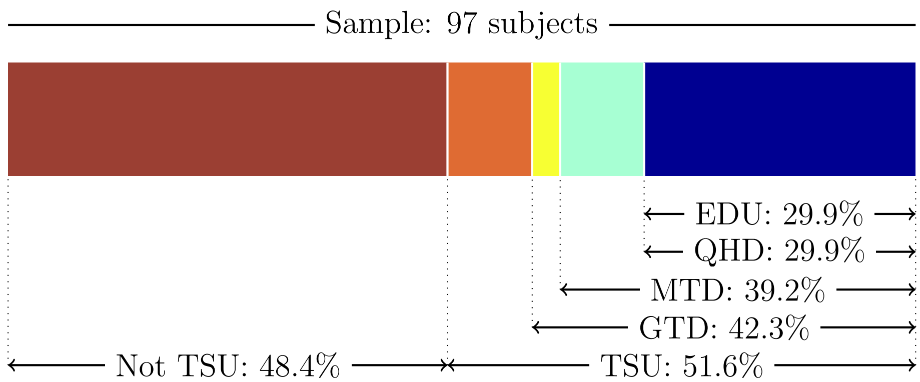


FIGURE 1: Classification of subjects in AS's experiment.

that the agent is M rational. For example, a subject is strictly QHD rational if her dataset passes the QHD test but not the EDU test.

Figure 1 summarizes the results. We find that 29 subjects are EDU rational. QHD also rationalizes the same 29 subjects: There are *no* subjects who are strictly QHD rational. As we mentioned before Proposition 1, this is related to agents' peculiar pattern of choices. Proposition 1 shows that if an agent does not consume at the soonest date (i.e., $x_0^k = 0$ for all $k \in K$), then EDU and PQHD are observationally equivalent. In AS's experiment, more than 82.8% of the subjects who satisfy SAR-EDU (i.e., 25% of the total subjects) do not consume at the soonest date.

EDU and QHD are arguably the most important models of intertemporal choice used in economics, but it is interesting to go beyond these models and look at the more general utility functions described in Section 4.2. We find that 9 additional subjects (9.3%) have utility functions in MTD, 3 additional subjects (3.1%) have utility functions in GTD, and 9 more subjects (9.3%) become rational by allowing a general TSU. In all, 51.6% of subjects can be rationalized by one of the time-separable models.

In summary, while AS find moderate support for EDU, our conclusion is closer to a rejection of EDU. In fact, close to half of the subjects in the experiment do not even exhibit time separable preferences. In the next section, we look at why our methods and AS's give seemingly contradictory conclusions from the same data.

Similarities and differences with AS's findings. Our analysis is for individual subjects. But the main results in AS use pooled data from all subjects. If we instead focus on individual level estimates of the same parametric model as AS, the source of the differences becomes quite clear. We focus on the individual estimates from AS (see Table A6-7 in the

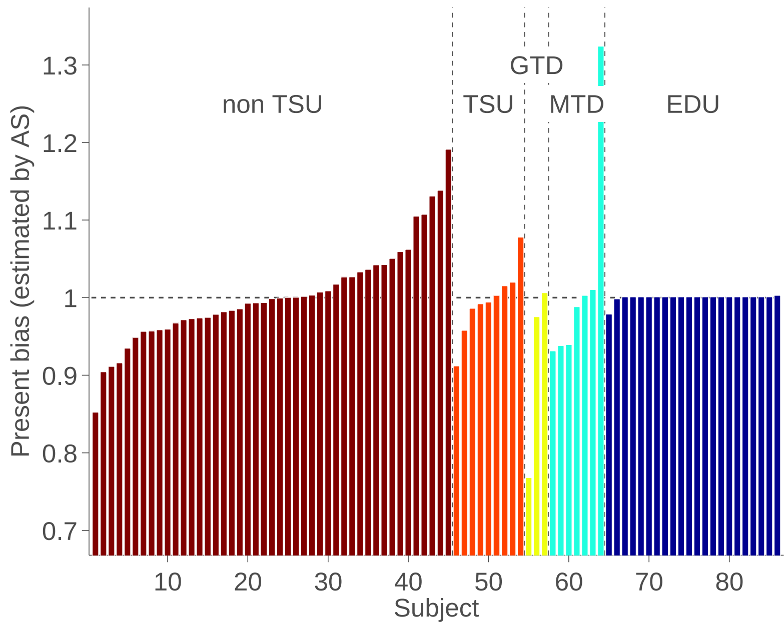


FIGURE 2: Estimated present-bias parameter for each category of subjects.

Online Appendix of Andreoni and Sprenger, 2012). [7]

Figure 2 summarizes the comparison. Each bar graph in the figure corresponds to one subject. The horizontal value of the bar is AS's estimated value $\widehat{\beta}$ for that subject. We categorized the subjects depending on their strict M-rationality. For example, the subjects who belong to the blue area pass our EDU test and the subjects who belong to the brown area do not pass any tests; they are not TSU rational.

There are two important facts one can glean from the figure. First, our test is consistent with AS's methodology and their estimates: the subjects who pass the EDU test have estimated $\widehat{\beta}$ very close to 1. So Figure 2 shows that our methodology and AS's methodology are, in fact, in agreement.

Secondly, those subjects who fail the EDU test but pass MTD, GTD, or TSU test tend to have $\widehat{\beta} \neq 1$. Moreover, those who do not pass any of the tests (i.e., non TSU subjects) have estimated $\widehat{\beta}$ which are far from 1 in magnitude compared to the other groups of subjects, and are distributed symmetrically around 1.[8] Roughly speaking, for half of the non-TSU subjects, $\widehat{\beta} > 1$; for the other half, $\widehat{\beta} < 1$. Hence, the "average" subject looks, in some sense,

[7]We obtain parameters for 86 of the 97 subjects. The remaining 11 subjects were excluded from AS's analysis, since preference parameters were not estimable. We can run our tests on the 11 excluded subjects: 7 of them pass the EDU and QHD tests.

[8]See the Online Appendix D for further comparisons between AS's parametric model estimation and our nonparametric revealed preference tests.

as an EDU agent, even though the majority of subjects are inconsistent with EDU according to our test. It is therefore possible that AS's finding in favor of EDU in their aggregate preference estimation reflects the choice behavior of such an average subject.

Choice pattern of EDU and non TSU subjects Next we look into subjects' choice patterns, focusing on the two main groups that the subjects fall into: those that are EDU rational and those that fail the TSU test. We investigate three aspects. First, we study the fraction of choices at the corner of the budget set. Second, we checked for violations of wealth monotonicity. Finally, we checked for violations of WARP.

Corner vs. interior choice: For each subject, we calculate the proportion (out of 45 choices) of (i) interior allocations, (ii) corner allocations in which subjects spend all their budget on a later reward (called "all tokens later"), and (iii) corner allocations in which subjects spend all their budget on the earlier reward (called "all tokens sooner").

We observe that all but two subjects who pass our EDU test never made interior allocations during the experiment, and frequently chose to allocate all tokens to the later payments. [9] This point is made clear in Figure 3, which presents each subject's choice pattern, sorted by the results of our EDU and TSU tests. The fraction of interior allocations increases by moving from EDU rational subjects (only 6.9% of them, 2 subjects, made at least one interior allocation) to subjects who pass the TSU test but not the EDU test (66.7% of them made at least one interior allocation); and it increases further when we look at subjects who fail the TSU test (in fact, 48.9% of them chose interior allocations in at least half of the trials).

The high incidence of corner solutions for EDU rational agents should be considered in light of Proposition 1. EDU and QHD are observationally equivalent when a subject never chooses date 0 consumption, and this happens for 82.8% (24 out of 29) of EDU rational subjects. So, for the vast majority of the subjects that pass SAR-EDU, the theory would have no power to distinguish between QHD and EDU.

Wealth Monotonicity: In AS's experiments, eight out of the nine time frames contain a wealth shift. We check wealth monotonicity, or normality of demand, using choices in those time frames. Monotonicity requires that c_τ and $c_{\tau+d}$ should be weakly increasing in wealth, holding the price rate constant in the eight time frames. Obviously, all of the EDU rational subjects satisfy monotonicity. On the other hand, most of non TSU subjects (43 out of 47 subjects) violate monotonicity.[10] So there is a simple explanation for why so many subject

[9]AS already remark on the incidence of of corner choices, and comment on how this phenomenon may suggest that the curvature of utility is small and close to that of a linear function.

[10]See Chakraborty et al. (2014) for a similar argument.

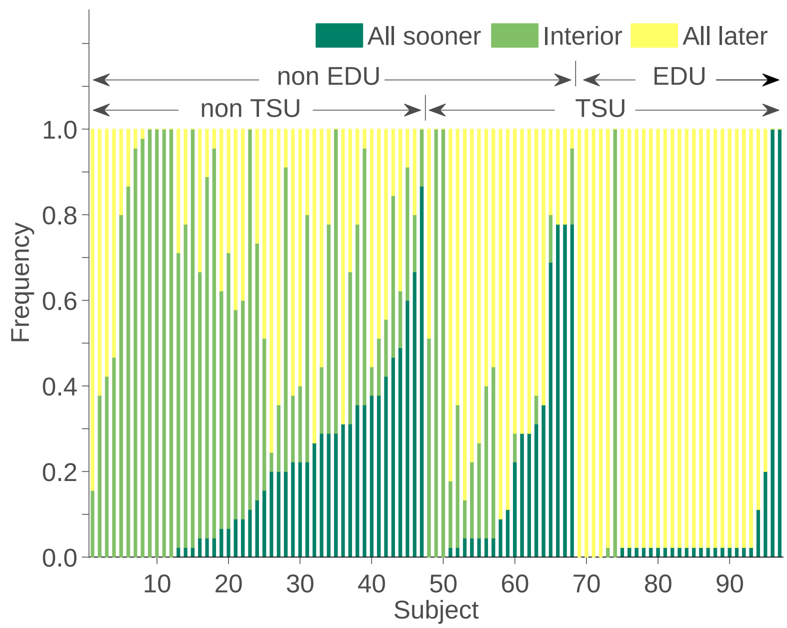


FIGURE 3: Individual choice patterns and class of rationality.

in AS's experiment are classified as non-TSU. They exhibit violations of normal demand.

Eliminating the wealth-shift observations from the data does not, however, suffice to make most subjects TSU rational. The choices of the non-TSU agents are inconsistent with TSU for more complicated reasons than simply a violation of normal demand.

WARP: In AS's experimental design, budget lines never cross at an interior point of the budget. However, there are budgets that cross at the corner of consuming all later (when "all later" corresponds to the same date; see Figure in the appendix) In particular, eight out of the nine time frames contain 4 budgets that share the same "all later" allocation at \$20. In the remaining time frame $(\tau, d) = (7, 70)$, all 5 budgets share the same "all later" allocation at \$20. So we can test WARP by using such choices. We found that 7 out of 97 subjects violated WARP. None of these 7 pass the TSU test (of course).

Distance measure. We find that many subjects' choices in the AS experiment are inconsistent with EDU, QHD, and even TSU. A natural question is then "how far" are these choices from EDU, QHD, or TSU rationality. Is a subject inconsistent with these models because she made a few mistakes, or is her behavior severely inconsistent with the model? The answer to these questions is that the violations we have detected are severe, and not due to small mistakes.

More precisely, to answer these questions we quantify the distance of the dataset from rationality by finding the largest subset of the dataset that passes the test under consideration. [11] In particular, we take the following steps. For each EDU non-rational (similarly for QHD and TSU) subject's dataset: (i) We randomly drop one observation from the dataset; (ii) We implement the EDU test. If the dataset is EDU rational, we stop here. Otherwise, we drop another observation randomly and test for EDU rationality again; and (iii) We repeat this procedure until the subset becomes EDU rational.

Ideally, one would check all possible subsets of data, but such a calculation is obviously computationally infeasible. Our approach of sequentially choosing (at random) one observation to drop is a rough approximation to the ideal measure. In particular, the conclusion can depend on the particular sequence chosen. To address this problem we iterate the process $30,000$ times for each EDU non-rational subject. [12] Let n_m be the number of observations required to be dropped from the original dataset to make the subdata EDU rational, in the m-th iteration. We define the distance of the dataset from EDU rationality by $d'_{\text{EDU}} = \min\{n_1, \ldots, n_{30000}\}/45$. By definition, the measure is between 0 and 1, and the smaller d'_{EDU} is the closer the dataset to be EDU rational. We also note that the measure is an upper bound on the distance we want to capture, due to the random nature and path-dependence of our approach. [13]

The left panel of Figure 4 shows the empirical CDFs of d'_{EDU} along with d'_{QHD} and d'_{TSU}. Note that the sample size is different for each line: d'_{EDU} and d'_{QHD} are calculated for the 68 EDU and QHD non-rational subjects, while d'_{TSU} is calculated for the 47 TSU non-rational subjects. We find that the median d'_{TSU} is 0.111, implying that half of the 47 TSU non-rational subjects become TSU rational by dropping at most 11% of the observations. For EDU and QHD, on the other hand, more observations need to be dropped to rationalize the data: median d'_{EDU} and d'_{QHD} are 0.378 and 0.356, respectively. This shows that subjects' violation of EDU and QHD are not due to small mistakes. [14]

In the right panel of Figure 4, we see a significant positive correlation (Pearson's corre-

[11]This approach is motivated by Houtman and Maks (1985), who measure the distance to rationality by finding the largest subset of observations that is consistent with GARP.

[12]We first performed 10,000 iterations and then prepared two additional sets, of 10,000 iterations each, as a way to check robustness of our approach. One might worry that this sampling approach may be far from the optimal exhaustive search over all subsets, but we increased the sample size very significantly without detecting important changes. We refer Section E of the Online Appendix for more details.

[13]We should observe $d'_{\text{EDU}} \geq d'_{\text{QHD}} \geq d'_{\text{TSU}}$ as a logical consequence (if the subset of data, after dropping n observations, is EDU rational, then the same subset is QHD rational, and so on). In reality, however, due to sample variations in the stochastic algorithm we use to compute distances, we observe several instances in which $d'_{\text{EDU}} \geq d'_{\text{QHD}}$ is violated. We correct for this by simply replacing d'_{QHD} with d'_{EDU} whenever such a violation is observed.

[14]We also find that the distributions of d'_{EDU} and d'_{QHD} are almost indistinguishable (the null hypothesis of equal distribution is not rejected by the two-sample Kolmogorov-Smirnov test).

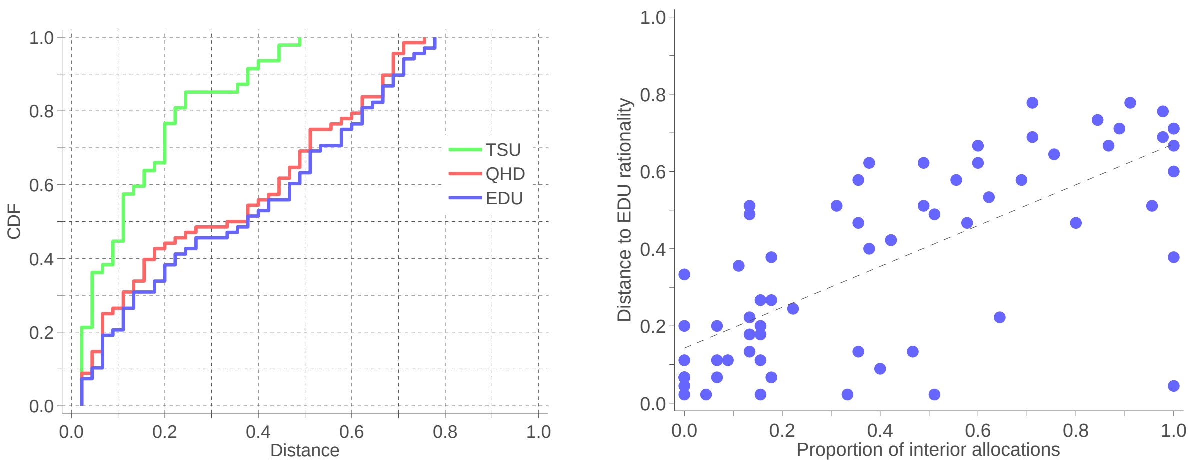


FIGURE 4: (Left) Empirical CDFs of distance measures of dataset from EDU, QHD, and TSU rationality. (Right) Distance to EDU rationality and proportion of interior allocations in the original dataset. The dotted line represents the slope of the least-squares fit.

lation coefficient $\rho = 0.7239$, $p < 10^{-11}$) between the proportion of interior allocations and the distance to EDU rationality. This correlation is in line with our speculation that the affluence of corner allocations make our revealed preference tests less demanding.

Jittering analysis. Aside from distance, we considered the robustness of our results in a different sense. We studied how "knife edge" the satisfaction of an axiom can be. Is it possible that subjects have preferences in model M, but that they have slightly unstable tastes? Could the violations of QHD be due to small instabilities in tastes? We employ a "jittering" method akin to the one discussed in Andreoni et al. (2013a). [15]

We perturb utility to produce data from a synthetic consumer with slightly unstable tastes: more precisely, we assume a CRRA instantaneous utility with QHD of the form (7), as in AS. Given a set of estimated parameters $(\widehat{\alpha}, \widehat{\delta}, \widehat{\beta})$, we added normal noise on one of the parameters while fixing the other two, e.g., $(\widehat{\alpha} + \varepsilon, \widehat{\delta}, \widehat{\beta})$ where $\varepsilon \sim N(0, \sigma^2)$. We set σ, the standard deviation of jittering, to equal the standard error of the corresponding parameter estimate. We simulate choices with such "jittered" parameters, and then apply our test.

First, we take parameters and standard errors from the aggregate estimation in AS: $(\alpha, \delta, \beta) = (0.897, 0.999, 1.007)$, $(\text{se}(\alpha), \text{se}(\delta), \text{se}(\beta)) = (0.0085, 1.8 \times 10^{-4}, 0.0058)$.[16] For each parameter, we simulate 1,000 jittered versions of parameters, predict choices, and perform the QHD test. We observe 100% pass rate no matter which parameter is jittered, suggesting

[15] We appreciate insightful comments from Jim Andreoni and Ben Gillen on this subject.

[16] Table 3, Column (3) in Andreoni and Sprenger (2012).

that our QHD test is robust to small perturbations to the underlying preference parameters.

Secondly, we perform the same exercise using AS's individual parameter estimates and standard errors, restricting our attention to those subjects who pass our QHD test (and whose parameters are estimable by AS). For each subject and each parameter, we draw 100 jittered versions of the parameter using estimated standard errors, predict choices, and perform the QHD test. This procedure gives us pass rates for QHD for each subject. We observe 100% pass rate for 20 out of 22 subjects when α is jittered, all 22 subjects when δ is jittered, and (iii) all 22 subjects when β is jittered. As in the case of the aggregate parameter estimates, the QHD test is robust to perturbation of the underlying preference parameters.

We have performed a similar exercise while perturbing choices instead of utility parameters. We prefer the method of perturbing utility because the story of slightly unstable tastes is more appealing than the idea that agents "tremble" when making a choice. The conclusion of this analysis is not as clearly in favor of the robustness of our tests, and it depends on what one takes to be the relevant jittering standard deviations. The results are in Section F of the Online Appendix.

Power of the tests. Finally, we discuss the power of our tests. It is well known that tests in revealed preference theory can have low power when used on certain configurations of budget sets. The low power of GARP is well documented. As a result, it is common to assess the power of a test by comparing the pass rates (the fraction of choices that pass the relevant revealed preference axiom) from purely random choices. [17] Here we report the results from such an assessment using our tests and the experimental design of AS. We find no evidence of low power.

We generate 10,000 datasets in which choices are made at random and uniformly distributed on the frontier of the budget set (Method 1 of Bronars, 1987). Datasets generated in this way always fail our tests (Table 1 shows pass rates). Next, we apply the simple bootstrap method to look at the power from an ex post perspective, as originally introduced in Andreoni and Miller (2002). For each of 45 budget sets, we randomly pick one choice from the set of choices observed in the entire experiment (i.e., 97 observations for each budget). We generate 10,000 such datasets and apply our revealed preference tests. We again observe high percentages of violation.

[17] The idea of using random choices as a benchmark is first applied to revealed preference theory by Bronars (1987). This approach is the most popular in empirical application: see, among other studies, Adams et al. (2014), Andreoni and Miller (2002), Beatty and Crawford (2011), Choi et al. (2007), Crawford (2010), Dean and Martin (forthcoming), Fisman et al. (2007). For overview of power calculation, see discussions in Andreoni et al. (2013a) and Crawford and De Rock (2014).

Table 1: Power measures.

Sampling	EDU	PQHD	QHD	MTD	GTD	TSU
Uniform random	0.00	0.00	0.00	0.00	0.00	0.00
Simple Bootstrap	0.00	0.00	0.00	0.00	0.00	0.00

The conclusion is that our tests seem to have good power against the (admittedly crude) alternative of random choices. This is a credit to the design of AS.

Afriat inequalities. It should be said that the empirical implementation of our test rests on a set of Afriat inequalities, and not on explicitly checking the axioms. The Afriat inequalities are new to our paper, though (see Lemma 1), and different from the standard approach to developing Afriat inequalities in the revealed preference literature. The new form of Afriat inequalities may seem ex-post (now that we know them) like a minor idea, but they were not ex-ante obvious. There are several papers (Crawford, 2010; Demuynck and Verriest, 2013; Adams et al., 2014) in the revealed preference literature that formulate the inequalities in the traditional fashion. The system of inequalities is then not linear (and cannot be linearized like our system can). As a result, these authors resort to a grid search over a finite set of values of the discount factor. The grid search can be a real limitation: We have examples in which our test gives higher pass rates for EDU than what the authors' methods give. Presumably the reason is that the grid does not allow one to conclude with certainty that an agent is not EDU rational, as it does not take full advantage of δ having arbitrary values in $(0, 1]$. So, in a sense, one of the key innovations of our paper are the new Afriat inequalities. These are crucial for both the theoretical results and the empirical implementation.[18]

6. Proof of Theorem 1

We present the proof of the equivalence between EDU rationality and SAR-EDU.

The proof is based on using the first-order conditions for maximizing a utility with the EDU over a budget set. Our first lemma ensures that we can without loss of generality restrict attention to first order conditions. The proof of the lemma is the same as that of Lemma 3 in Echenique and Saito (forthcoming) with the changes of T to S and $\{\delta^t\}_{t\in T}$ to $\{\mu_s\}_{s\in S}$, where μ_s is the subjective probability that state s realizes.

[18]The other key theoretical insight is the approximation result in Lemma 6.

We use the following notation in the proofs:

$$\mathcal{X} = \{x_t^k : k \in K, t \in T\}.$$

Lemma 1. *Let $(x^k, p^k)_{k=1}^K$ be a dataset. The following statements are equivalent:*

1. *$(x^k, p^k)_{k=1}^K$ is EDU rational.*

2. *There are strictly positive numbers v_t^k, λ^k, and $\delta \in (0, 1]$, for $t = 1, \ldots, T$ and $k = 1, \ldots, K$, such that*

$$\delta^t v_t^k = \lambda^k p_t^k, \quad x_t^k > x_{t'}^{k'} \implies v_t^k \leq v_{t'}^{k'}.$$

Proof. We shall prove that (1) implies (2). Let $(x^k, p^k)_{k=1}^K$ be EDU rational. Let $\delta \in (0, 1]$ and $u : \mathbf{R}_+ \to \mathbf{R}$ be as in the definition of EDU rational data. Then (see, for example, Theorem 28.3 of Rockafellar, 1997), there are numbers $\lambda^k \geq 0$, $k = 1, \ldots, K$ such that if we let

$$v_t^k = \frac{\lambda^k p_t^k}{\delta^t}$$

then $v_t^k \in \partial u(x_t^k)$ if $x_t^k > 0$, and there is $\underline{w} \in \partial u(x_t^k)$ with $v_t^k \geq \underline{w}$ if $x_t^k = 0$. In fact, it is easy to see that $\lambda^k > 0$, and therefore $v_t^k > 0$.

By the concavity of u, and the consequent monotonicity of $\partial u(x_t^k)$ (see Theorem 24.8 of Rockafellar, 1997), if $x_t^k > x_{t'}^{k'} > 0$, $v_t^k \in \partial u(x_t^k)$, and $v_{t'}^{k'} \in \partial u(x_{t'}^{k'})$, then $v_t^k \leq v_{t'}^{k'}$. If $x_t^k > x_{t'}^{k'} = 0$, then $\underline{w} \in \partial u(x_{t'}^{k'})$ with $v_{t'}^{k'} \geq \underline{w}$. So $v_t^k \leq \underline{w} \leq v_{t'}^{k'}$.

In second place, we show that (2) implies (1). Suppose that the numbers v_t^k, λ^k, δ, for $t \in T$ and $k \in K$, are as in (2).

Enumerate the elements in $\mathcal{X}$ in increasing order:

$$y_1 < y_2 < \ldots < y_n.$$

Let

$$\underline{y}_i = \min\{v_t^k : x_t^k = y_i\} \text{ and } \bar{y}_i = \max\{v_t^k : x_t^k = y_i\}.$$

Let $z_i = (y_i + y_{i+1})/2$, $i = 1, \ldots, n-1$; $z_0 = 0$, and $z_n = y_n + 1$. Let f be a correspondence defined as follows:

$$f(z) = \begin{cases} [\underline{y}_i, \bar{y}_i] & \text{if } z = y_i, \\ \max\{\bar{y}_i : z < y_i\} & \text{if } y_n > z \text{ and } \forall i (z \neq y_i), \\ \underline{y}_n/2 & \text{if } y_n < z. \end{cases}$$

By assumption of the numbers v_t^k, we have that, when $y < y'$, $v \in f(y)$ and $v' \in f(y')$, then $v \leq v'$. Then the correspondence f is monotone and there is a concave function u for which $\partial u = f$ (Theorem 24.8 of Rockafellar, 1997). Given that $v_t^k > 0$ all the elements in the range of f are positive, and therefore u is strictly increasing.

Finally, for all (k,t), $\lambda^k p_t^k / \delta^t = v_t^k \in \partial u(v_t^k)$ and therefore the first-order conditions to a maximum choice of x hold at x_t^k. Since u is concave the first-order conditions are sufficient. The dataset is therefore EDU rational. □

6.1. Necessity

Lemma 2. *If a dataset $(x^k, p^k)_{k=1}^K$ is EDU rational, then it satisfies SAR-EDU.*

Proof. Let $(x^k, p^k)_{k=1}^K$ be EDU rational, and let $\delta \in (0,1]$ and $u : \mathbf{R}_+ \to \mathbf{R}$ be as in the definition of EDU rational. By Lemma 1, there exists a strictly positive solution v_t^k, λ^k, δ to the system in Statement (2) of Lemma 1 with $v_t^k \in \partial u(x_t^k)$ when $x_t^k > 0$, and $v_t^k \geq \underline{w} \in \partial u(x_t^k)$ when $x_t^k = 0$.

Let $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$ be a sequence satisfying the three conditions in SAR-EDU. Then $x_{t_i}^{k_i} > x_{t_i'}^{k_i'}$. Suppose that $x_{t_i'}^{k_i'} > 0$. Then, $v_{t_i}^{k_i} \in \partial u(x_{t_i}^{k_i})$ and $v_{t_i'}^{k_i'} \in \partial u(x_{t_i'}^{k_i'})$. By the concavity of u, it follows that $\lambda^{k_i} \delta^{t_i'} p_{t_i}^{k_i} \leq \lambda^{k_i'} \delta^{t_i} p_{t_i'}^{k_i'}$ (see Theorem 24.8 of Rockafellar, 1997). Similarly, if $x_{t_i'}^{k_i'} = 0$, then $v_{t_i}^{k_i} \in \partial u(x_{t_i}^{k_i})$ and $v_{t_i'}^{k_i'} \geq \underline{w} \in \partial u(x_{t_i'}^{k_i'})$. So $\lambda^{k_i} \delta^{t_i'} p_{t_i}^{k_i} \leq \lambda^{k_i'} \delta^{t_i} p_{t_i'}^{k_i'}$ Therefore,

$$1 \geq \prod_{i=1}^n \frac{\lambda^{k_i} \delta^{t_i'} p_{t_i}^{k_i}}{\lambda^{k_i'} \delta^{t_i} p_{t_i'}^{k_i'}} = \frac{1}{\delta^{(\sum t_i - \sum t_i')}} \prod_{i=1}^n \frac{p_{t_i}^{k_i}}{p_{t_i'}^{k_i'}} \geq \prod_{i=1}^n \frac{p_{t_i}^{k_i}}{p_{t_i'}^{k_i'}},$$

as the sequence satisfies (2) and (3) of SAR-EDU; and hence $\sum t_i \geq \sum t_i'$ and the numbers λ^k appear the same number of times in the denominator as in the numerator of this product. □

6.2. Theorem of the Alternative

To prove sufficiency, we shall use the following lemma, which is a version of the Theorem of the Alternative. This is Theorem 1.6.1 in Stoer and Witzgall (1970). We shall use it here in the cases where F is either the real or the rational numbers.

Lemma 3. *Let A be an $m \times n$ matrix, B be an $l \times n$ matrix, and E be an $r \times n$ matrix. Suppose that the entries of the matrices A, B, and E belong the a commutative ordered field $\mathbf{F}$. Exactly one of the following alternatives is true.*

1. *There is $u \in \mathbf{F}^n$ such that $A \cdot u = 0$, $B \cdot u \geq 0$, $E \cdot u \gg 0$.*

2. *There is* $\theta \in \mathbf{F}^r$, $\eta \in \mathbf{F}^l$, *and* $\pi \in \mathbf{F}^m$ *such that* $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$; $\pi > 0$ *and* $\eta \geq 0$.

We also use the following lemma, which follows from Lemma 3 (See Border (2013) or Chambers and Echenique (2014)):

Lemma 4. *Let* A *be an* $m \times n$ *matrix,* B *be an* $l \times n$ *matrix, and* E *be an* $r \times n$ *matrix. Suppose that the entries of the matrices* A, B, *and* E *are rational numbers. Exactly one of the following alternatives is true.*

1. *There is* $u \in \mathbf{R}^n$ *such that* $A \cdot u = 0$, $B \cdot u \geq 0$, *and* $E \cdot u \gg 0$.
2. *There is* $\theta \in \mathbf{Q}^r$, $\eta \in \mathbf{Q}^l$, *and* $\pi \in \mathbf{Q}^m$ *such that* $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$; $\pi > 0$ *and* $\eta \geq 0$.

6.3. Sufficiency

We proceed to prove the sufficiency direction. An outline of the argument is as follows. We know from Lemma 1 that it suffices to find a solution to the Afriat inequalities (actually first order conditions), written as statement (2) in the lemma. So we set up the problem to find a solution to a system of linear inequalities obtained from using logarithms to linearize the Afriat inequalities in Lemma 1.

Lemma 5 establishes that SAR-EDU is sufficient for SEU rationality when the logarithms of the prices are rational numbers. The role of rational logarithms comes from our use of a version the theorem of the alternative (see Lemma 4).

The next step in the proof (Lemma 6) establishes that we can approximate any dataset satisfying SAR-EDU with a dataset for which the logarithms of prices are rational, and for which SAR-EDU is satisfied. This step is crucial, and somewhat delicate.[19]

Finally, Lemma 7 establishes the result by using another version of the theorem of the alternative, stated as Lemma 3 above.

The statement of the lemmas follow. The rest of the paper is devoted to the proof of these lemmas.

Lemma 5. *Let data* $(x^k, p^k)_{k=1}^k$ *satisfy SAR-EDU. Suppose that* $\log(p_t^k) \in \mathbf{Q}$ *for all* k *and* t. *Then there are numbers* v_t^k, λ^k, δ, *for* $t \in T$ *and* $k = 1, \ldots, K$ *satisfying* (2) *in Lemma 1.*

[19] One might have tried to obtain a solution to the Afriat inequalities for "perturbed" systems (with prices that are rational after taking logs), and then considered the limit. This does not work because the solutions to our systems of inequalities are in a non-compact space. It is not clear how to establish that the limits exist and are well-behaved. Lemma 6 avoids the problem.

Lemma 6. *Let data* $(x^k, p^k)_{k=1}^k$ *satisfy SAR-EDU. Then for all positive numbers* $\overline{\varepsilon}$*, there exists* $q_t^k \in [p_t^k - \overline{\varepsilon}, p_t^k]$ *for all* $t \in T$ *and* $k \in K$ *such that* $\log q_t^k \in \mathbf{Q}$ *and the dataset* $(x^k, q^k)_{k=1}^k$ *satisfy SAR-EDU.*

Lemma 7. *Let data* $(x^k, p^k)_{k=1}^k$ *satisfy SAR-EDU. Then there are numbers* v_t^k*,* λ^k*,* δ*, for* $t \in T$ *and* $k = 1, \ldots, K$ *satisfying* (2) *in Lemma 1.*

6.4. *Proof of Lemma 5*

We linearize the equation in System (2) of Lemma 1. The result is:

$$\log v(x_t^k) + t \log \delta - \log \lambda^k - \log p_t^k = 0, \tag{8}$$

$$x > x' \Longrightarrow \log v(x') \geq \log v(x), \tag{9}$$

$$\log \delta \leq 0. \tag{10}$$

In the system comprised by (8), (9), and (10), the unknowns are the real numbers $\log v_t^k$, $\log \delta$, $k \in K$ and $t \in T$.

First, we are going to write the system of inequalities (8) and (9) in matrix form. We shall define a matrix A such that there are positive numbers v_t^k, λ^k, δ the logs of which satisfy Equation (8) if and only if there is a solution $u \in \mathbf{R}^{K \times (T+1)+1+K+1}$ to the system of equations

$$A \cdot u = 0,$$

and for which the last component of u is strictly positive.

Let A be a matrix with $K \times (T + 1) + 1 + K + 1$ columns, defined as follows: We have one row for every pair (k, t); one column for every pair (k, t); one column for each k; and two additional columns. Organize the columns so that we first have the $K \times (T + 1)$ columns for the pairs (k, t); then one of the single columns mentioned in last place, which we shall refer to as the δ-column; then K columns (one for each k); and finally one last column. In the row corresponding to (k, t) the matrix has zeroes everywhere with the following exceptions: it has a 1 in the column for (k, t); it has a t in the δ column; it has a -1 in the column for k; and $-\log p_t^k$ in the very last column.

Thus, matrix A looks as follows:

$$\begin{array}{c} \\ \vdots \\ {\scriptstyle (k,t)} \\ \vdots \end{array}
\begin{array}{c}
\begin{array}{ccccccccccccc} {\scriptstyle (1,0)} & \cdots & {\scriptstyle (k,t)} & \cdots & {\scriptstyle (K,T)} & {\scriptstyle \delta} & {\scriptstyle 1} & \cdots & {\scriptstyle k} & \cdots & {\scriptstyle K} & {\scriptstyle p} \end{array} \\
\left[\begin{array}{ccccc|c|ccccc|c}
\vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & 1 & \cdots & 0 & t & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\
\vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots
\end{array}\right]
\end{array}$$

Consider the system $A \cdot u = 0$. If there are numbers solving Equation (8), then these define a solution $u \in \mathbf{R}^{K\times(T+1)+1+K+1}$ for which the last component is 1. If, on the other hand, there is a solution $u \in \mathbf{R}^{K\times(T+1)+1+K+1}$ to the system $A \cdot u = 0$ in which the last component is strictly positive, then by dividing through by the last component of u we obtain numbers that solve Equation (8).

In second place, we write the system of inequalities (9) and (10) in matrix form. Let B be a matrix with $K \times (T+1)+1+K+1$ columns. Define B as follows: One row for every pair (k,t) and (k',t') with $x_t^k > x_{t'}^{k'}$; in the row corresponding to (k,t) and (k',t') we have zeroes everywhere with the exception of a -1 in the column for (k,t) and a 1 in the column for (k',t') . These rows captures the inequality (9). Finally, in the last row, we have zero everywhere with the exception of a -1 at $K \times (T+1)+1$th column. We shall refer to this last row as the δ-row, which capturing the inequality (10).

In third place, we have a matrix E that captures the requirement that the last component of a solution be strictly positive. The matrix E has a single row and $K \times (T+1)+1+K+1$ columns. It has zeroes everywhere except for 1 in the last column.

To sum up, there is a solution to system (8), (9) and (10) if and only if there is a vector $u \in \mathbf{R}^{K\times(T+1)+1+K+1}$ that solves the system of equations and linear inequalities

$$(S1): \ A \cdot u = 0, \ B \cdot u \geq 0, \ E \cdot u \gg 0.$$

The entries of A, B, and E are integer numbers, with the exception of the last column of A. Under the hypothesis of the lemma we are proving, the last column consists of rational numbers.

By Lemma 4, then, there is such a solution u to $S1$ if and only if there is no rational vector (θ, η, π) that solves the system of equations and linear inequalities

$$(S2): \ \theta \cdot A + \eta \cdot B + \pi \cdot E = 0, \ \eta \geq 0, \ \pi > 0.$$

In the following, we shall prove that the non-existence of a solution u implies that the data must violate SAR-EDU. Suppose then that there is no solution u and let (θ, η, π) be a rational vector as above, solving system $S2$.

By multiplying (θ, η, π) by any positive integer we obtain new vectors that solve $S2$, so we can take (θ, η, π) to be integer vectors.

Henceforth, we use the following notational convention: For a matrix D with $K \times (T+1)+1+K+1$ columns, write D_1 for the submatrix of D corresponding to the first $K \times (T+1)$ columns; let D_2 be the submatrix corresponding to the following one column (i.e., δ-column); D_3 correspond to the next K columns; and D_4 to the last column. Thus, $D = [D_1|D_2|D_3|D_4]$.

Claim 1. *(i) $\theta \cdot A_1 + \eta \cdot B_1 = 0$; (ii) $\theta \cdot A_2 + \eta \cdot B_2 = 0$; (iii) $\theta \cdot A_3 = 0$; and (iv) $\theta \cdot A_4 + \pi \cdot E_4 = 0$.*

Proof. Since $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$, then $\theta \cdot A_i + \eta \cdot B_i + \pi \cdot E_i = 0$ for all $i = 1, \ldots, 4$. Moreover, since B_3, B_4, E_1, E_2, and E_3 are zero matrices, we obtain the claim. □

For convenience, we transform the matrices A and B using θ and η. We transform the matrices A, and B as follows. Let us define a matrix A^* from A by letting A^* have $K \times (T+1) + 1 + K + 1$ columns that consists of the rows as follows: for each row in $r \in A$ (i) have θ_r copies of the rth row when $\theta_r > 0$; (ii) omit row r when $\theta_r = 0$; and (iii) have θ_r copies of the rth row multiplied by -1 when $\theta_r < 0$.

We refer to rows that are copies of some r in A with $\theta_r > 0$ as *original* rows. We refer to rows that are copies of some r in A with $\theta_r < 0$ as *converted* rows.

Similarly, we define the matrix B^* from B by including the same columns as B and η_r copies of each row (and thus omitting row r when $\eta_r = 0$; recall that $\eta_r \geq 0$ for all r).

Claim 2. *For any (k,t), all the entries in the column for (k,t) in A^*_1 are of the same sign.*

Proof. By definition of A, the column for (k,t) will have zero in all its entries with the exception of the row for (k,t). In A^*, for each (k,t), there are three mutually exclusive possibilities: the row for (k,t) in A can (i) not appear in A^*, (ii) it can appear as original, or (iii) it can appear as converted. This shows the claim. □

Claim 3. *There exists a sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ that satisfies Condition (1) in SAR-EDU.*

Proof. We define such a sequence by induction. Let $B^1 = B^*$. Given B^i, define B^{i+1} as follows.

Denote by $>^i$ the binary relation on $\mathcal{X}$ defined by $z >^i z'$ if $z > z'$ and there is at least one pair (k,t) and (k',t') for which (i) $x^k_t > x^{k'}_{t'}$; (ii) $z = x^k_t$ and $z' = x^{k'}_{t'}$; and (iii) the row corresponding $x^k_t > x^{k'}_{t'}$ in B has strictly positive weight in η.

The binary relation $>^i$ cannot exhibit cycles because $>^i \subseteq >$. There is therefore at least one sequence $z^i_1, \ldots z^i_{L_i}$ in $\mathcal{X}$ such that $z^i_j >^i z^i_{j+1}$ for all $j = 1, \ldots, L_i - 1$ and with the property that there is no $z \in \mathcal{X}$ with $z >^i z^i_1$ or $z^i_{L_i} >^i z$.

Observe that B^i has at least one row corresponding to $z^i_j >^i z^i_{j+1}$ for all $j = 1, \ldots, L_i - 1$. Let the matrix B^{i+1} be defined as the matrix obtained from B^i by omitting one copy of the row corresponding to $z^i_j > z^i_{j+1}$, for all $j = 1, \ldots L_i - 1$.

The matrix B^{i+1} has strictly fewer rows than B^i. There is therefore n^* for which B^{n^*+1} either has no more rows, or $B^{n^*+1}_1$ has only zeroes in all its entries (its rows are copies of the δ-row which has only zeroes in its first $K \times (T+1)$ columns).

Define a sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ by letting $x^{k_i}_{t_i} = z^i_1$ and $x^{k'_i}_{t'_i} = z^i_{L_i}$. Note that, as a result, $x^{k_i}_{t_i} > x^{k'_i}_{t'_i}$ for all i. Therefore the sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ satisfies Condition (1) in SAR-EDU. □

We shall use the sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ as our candidate violation of SAR-EDU.

Consider a sequence of matrices A^i, $i = 1, \ldots, n^*$ defined as follows. Let $A^1 = A^*$, $B^1 = B^*$, and $C^1 = \begin{bmatrix} A^1 \\ B^1 \end{bmatrix}$. Observe that the rows of C^1 add to the null vector by Claim 1.

We shall proceed by induction. Suppose that A^i has been defined, and that the rows of $C^i = \begin{bmatrix} A^i \\ B^i \end{bmatrix}$ add to the null vector.

Recall the definition of the sequence

$$x^{k_i}_{t_i} = z^i_1 > \ldots > z^i_{L_i} = x^{k'_i}_{t'_i}.$$

There is no $z \in \mathcal{X}$ with $z >^i z^i_1$ or $z^i_{L_i} >^i z$, so in order for the rows of C^i to add to zero there must be a -1 in A^i_1 in the column corresponding to (k'_i, t'_i) and a 1 in A^i_1 in the column corresponding to (k_i, t_i). Let r_i be a row in A^i corresponding to (k_i, t_i), and r'_i be a row corresponding to (k'_i, t'_i). The existence of a -1 in A^i_1 in the column corresponding to (k'_i, t'_i), and a 1 in A^i_1 in the column corresponding to (k_i, t_i), ensures that r_i and r'_i exist. Note that the row r'_i is a converted row while r_i is original. Let A^{i+1} be defined from A^i by deleting the two rows, r_i and r'_i.

Claim 4. *The sum of r_i, r'_i, and the rows of B^i which are deleted when forming B^{i+1} (corresponding to the pairs $z^i_j > z^i_{j+1}$, $j = 1, \ldots, L_i - 1$) add to the null vector.*

Proof. Recall that $z^i_j >^i z^i_{j+1}$ for all $j = 1, \ldots, L_i - 1$. So when we add the rows corresponding to $z^i_j >^i z^i_{j+1}$ and $z^i_{j+1} >^i z^i_{j+2}$, then the entries in the column for (k, t) with $x^k_t = z^i_{j+1}$ cancel out and the sum is zero in that entry. Thus, when we add the rows of B^i that are not in B^{i+1} we obtain a vector that is 0 everywhere except the columns corresponding to z^i_1 and $z^i_{L_i}$. This vector cancels out with $r_i + r'_i$, by definition of r_i and r'_i. □

Claim 5. *The matrix A^* can be partitioned into pairs (r_i, r'_i), in which the rows r'_i are converted and the rows r_i are original.*

Proof. For each i, A^{i+1} differs from A^i in that the rows r_i and r'_i are removed from A^i to form A^{i+1}. We shall prove that A^* is composed of the $2n^*$ rows r_i, r'_i.

First note that since the rows of C^i add up to the null vector, and A^{i+1} and B^{i+1} are obtained from A^i and B^i by removing a collection of rows that add up to zero, then the rows of C^{i+1} must add up to zero as well.

By way of contradiction, suppose that there exist rows left after removing r_{n^*} and r'_{n^*}. Then, by the argument above, the rows of the matrix C^{n^*+1} must add to the null vector. If there are rows left, then the matrix C^{n^*+1} is well defined.

By definition of the sequence B^i, however, B^{n^*+1} has all its entries equal to zero, or has no rows. Hence, the rows remaining in $A_1^{n^*+1}$ must add up to zero. By Claim 2, the entries of a column (k,t) of A^* are always of the same sign. Moreover, each row of A^* has a non-zero element in the first $K \times S$ columns. Therefore, no subset of the columns of A_1^* can sum to the null vector. $\square$

Claim 6. *(i) For any k and t, if $(k_i, t_i) = (k, t)$ for some i, then the row r_i corresponding to (k, t) appears as original in A^*. Similarly, if $(k'_i, t'_i) = (k', t')$ for some i, then the row corresponding to (k, t) appears converted in A^*.*
(ii) If the row corresponding to (k, t) appears as original in A^*, then there is some i with $(k_i, t_i) = (k, t)$. Similarly, if the row corresponding to (k, t) appears converted in A^*, then there is i with $(k'_i, t'_i) = (k, t)$.

Proof. (i) is true by definition of $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})$. (ii) is immediate from Claim 5 because if the row corresponding to (k, t) appears original in A^* then it equals r_i for some i, and then $x^k_t = x^{k_i}_{t_i}$. Similarly when the row appears converted. $\square$

Claim 7. *The sequence $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})_{i=1}^{n^*}$ satisfies Condition (2) and (3) in SAR-EDU.*

Proof. We first establish Condition (2). Note that A_2^* is a vector, and in row r the entry of A_2^* is as follows. There must be a raw (k, t) in A of which the raw r is a copy. Hence, the component at the row r of A_2^* is t if r is original and $-t$ if r is converted. Now, by the construction of the sequence when r appears as original there is some i for which $t = t_i$, when r appears as converted there is some i for which $t = t'_i$. So for each r there is i such that $(A_4^*)_r$ is either t_i or $-t'_i$. By Claim 1 (ii), $\theta \cdot A_2 + \eta \cdot B_2 = 0$. Recall that $\theta \cdot A_2$ equals the sum of the rows of A_2^*. Moreover, B_2 is a vector that has zeroes everywhere except a -1 in the δ row (i.e., $K \times (T+1) + 1$th row). Therefore, the sum of the rows of A_2^* equals $\eta_{K\times(T+1)+1}$, where $\eta_{K\times(T+1)+1}$ is the $K \times (T+1) + 1$th element of η. Since $\eta \geq 0$, therefore, $\sum_{i=1}^{n^*} t_i \geq \sum_{i=1}^{n^*} t'_i$, and Condition (2) in the axiom is satisfied.

Now we turn to (3). By Claim 1 (iii), the rows of A_3^* add up to zero. Therefore, the number of times that k appears in an original row equals the number of times that it appears in a converted row. By Claim 6, then, the number of times k appears as k_i equals the number of times it appears as k_i'. Therefore, Condition (3) in the axiom is satisfied. □

Finally, in the following, we show that $\prod_{i=1}^{n^*} p_{t_i}^{k_i}/p_{t_i'}^{k_i'} > 1$, which finishes the proof of Lemma 5 as the sequence $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^{n^*}$ would then exhibit a violation of SAR-EDU.

Claim 8. $\prod_{i=1}^{n^*} p_{t_i}^{k_i}/p_{t_i'}^{k_i'} > 1$.

Proof. By Claim 1 (iv) and the fact that the submatrix E_4 equals the scalar 1, we obtain

$$0 = \theta \cdot A_4 + \pi E_4 = \left(\sum_{i=1}^{n^*} (r_i + r_i') \right)_4 + \pi,$$

where $(\sum_{i=1}^{n^*}(r_i + r_i'))_4$ is the (scalar) sum of the entries of A_4^*. Recall that $-\log p_{t_i}^{k_i}$ is the last entry of row r_i and that $\log p_{t_i'}^{k_i'}$ is the last entry of row r_i', as r_i' is converted and r_i original. Therefore the sum of the rows of A_4^* are $\sum_{i=1}^{n^*} \log(p_{t_i'}^{k_i'}/p_{t_i}^{k_i})$. Then,

$$\sum_{i=1}^{n^*} \log(p_{t_i'}^{k_i'}/p_{t_i}^{k_i}) = -\pi < 0.$$

Thus $\prod_{i=1}^{n^*} p_{t_i}^{k_i}/p_{t_i'}^{k_i'} > 1$. □

6.5. *Proof of Lemma 6*

For each sequence $\sigma = (x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^{n}$ that satisfies conditions (1), (2), and (3) in SAR-EDU, we define a vector $t_\sigma \in \mathbf{N}^{(K \times T)^2}$ as follows. To make the notation easier, we identify the pair $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})$ with $((k_i, t_i), (k_i', t_i'))$. Let $t_\sigma((k,t),(k',t'))$ be the number of times that the pair $(x_t^k, x_{t'}^{k'})$ appears in the sequence σ. One can then describe the satisfaction of SAR-EDU by means of the vectors t_σ. Define

$$T = \left\{ t_\sigma \in \mathbf{N}^{(K \times T)^2} : \sigma \text{ satisfies Conditions (1), (2), (3) in SAR-EDU} \right\}.$$

Observe that the set T depends only on $(x^k)_{k=1}^K$ in the dataset $(x^k, p^k)_{k=1}^K$. It does not depend on prices.

For each $((k,t),(k',t')) \in (K \times T)^2$ such that $x_t^k > x_{t'}^{k'}$, define

$$\widehat{\gamma}((k,t),(k',t')) = \log \left(\frac{p_t^k}{p_{t'}^{k'}} \right),$$

and define $\widehat{\gamma}((k,t),(k',t')) = 0$ when $x^k_t \leq x^{k'}_{t'}$. Then, $\widehat{\gamma}$ is a $(KT)^2$-dimensional real-valued vector. If $\sigma = (x^{k_i}_{t_i}, x^{k'_i}_{t'_i})_{i=1}^n$, then

$$\widehat{\gamma} \cdot t_\sigma = \sum_{((k,t),(k',t')) \in (K \times T)^2} \widehat{\gamma}((k,t),(k',t')) t_\sigma((k,t),(k',t')) = \log \left(\prod_{i=1}^n \frac{p^{k_i}_{t_i}}{p^{k'_i}_{t'_i}} \right).$$

So the data satisfy SAR-EDU if and only if $\hat{\gamma} \cdot t \leq 0$ for all $t \in T$.

Enumerate the elements in $\mathcal{X}$ in increasing order:

$$y_1 < y_2 < \cdots < y_N.$$

Fix an arbitrary $\underline{\xi} \in (0,1)$.

We shall construct by induction a sequence $(\varepsilon^k_t(n))$ for $n = 1, \ldots, N$, where $\varepsilon^k_t(n)$ is defined for all (k,t) with $x^k_t = y_n$.

By the denseness of the rational numbers, and the continuity of the exponential function, for each (k,t) such that $x^k_t = y_1$, there exists a positive number $\varepsilon^k_t(1)$ such that $\log(p^k_t \varepsilon^k_t(1)) \in \mathbf{Q}$ and $\underline{\xi} < \varepsilon^k_t(1) < 1$. Let $\varepsilon(1) = \min\{\varepsilon^k_t(1) : x^k_t = y_1\}$.

In second place, for each (k,t) such that $x^k_t = y_2$, there exists a positive $\varepsilon^k_t(2)$ such that $\log(p^k_t \varepsilon^k_t(2)) \in \mathbf{Q}$ and $\underline{\xi} < \varepsilon^k_t(2) < \varepsilon(1)$. Let $\varepsilon(2) = \min\{\varepsilon^k_t(2) : x^k_t = y_2\}$.

In third place, and reasoning by induction, suppose that $\varepsilon(n)$ has been defined and that $\underline{\xi} < \varepsilon(n)$. For each (k,t) such that $x^k_t = y_{n+1}$, let $\varepsilon^k_t(n+1) > 0$ be such that $\log(p^k_t \varepsilon^k_t(n+1)) \in \mathbf{Q}$, and $\underline{\xi} < \varepsilon^k_t(n+1) < \varepsilon(n)$. Let $\varepsilon(n+1) = \min\{\varepsilon^k_t(n+1) : x^k_t = y_n\}$.

This defines the sequence $(\varepsilon^k_t(n))$ by induction. Note that $\varepsilon^k_t(n+1)/\varepsilon(n) < 1$ for all n. Let $\bar{\xi} < 1$ be such that $\varepsilon^k_t(n+1)/\varepsilon(n) < \bar{\xi}$.

For each $k \in K$ and $t \in T$, let $q^k_t = p^k_t \varepsilon^k_t(n)$, where n is such that $x^k_t = y_n$. We claim that the data $(x^k, q^k)_{k=1}^K$ satisfy SAR-EDU. Let γ^* be defined from $(q^k)_{k=1}^K$ in the same manner as $\widehat{\gamma}$ was defined from $(p^k)_{k=1}^K$.

For each pair $((k,t),(k',t'))$ with $x^k_t > x^{k'}_{t'}$, if n and m are such that $x^k_t = y_n$ and $x^{k'}_{t'} = y_m$, then $n > m$. By the definition of ε,

$$\frac{\varepsilon^k_t(n)}{\varepsilon^{k'}_{t'}(m)} < \frac{\varepsilon^k_t(n)}{\varepsilon(m)} < \bar{\xi} < 1.$$

Hence,

$$\gamma^*((k,t),(k',t')) = \log \frac{p^k_t \varepsilon^k_t(n)}{p^{k'}_{t'} \varepsilon^{k'}_{t'}(m)} < \log \frac{p^k_t}{p^{k'}_{t'}} + \log \bar{\xi} < \log \frac{p^k_t}{p^{k'}_{t'}} = \widehat{\gamma}(x^k_s, x^{k'}_{t'}).$$

Thus, for all $t \in T$, $\gamma^* \cdot t \leq \widehat{\gamma} \cdot t \leq 0$, as $t \geq 0$ and the data $(x^k, p^k)_{k=1}^K$ satisfy SAR-EDU. Thus the data $(x^k, q^k)_{k=1}^K$ satisfy SAR-EDU. Finally, note that $\underline{\xi} < \varepsilon^k_t(n) < 1$ for all n and each $k \in K, t \in T$. So that by choosing $\underline{\xi}$ close enough to 1 we can take the prices (q^k) to be as close to (p^k) as desired.

6.6. *Proof of Lemma 7*

Consider the system comprised by (8), (9), and (10) in the proof of Lemma 5. Let A, B, and E be constructed from the data as in the proof of Lemma 5. The difference with respect to Lemma 5 is that now the entries of A_4 may not be rational. Note that the entries of E, B, and A_i, $i = 1, 2, 3$ are rational.

Suppose, towards a contradiction, that there is no solution to the system comprised by (8), (9), and (10). Then, by the argument in the proof of Lemma 5 there is no solution to System $S1$. Lemma 3 with $\mathbf{F} = \mathbf{R}$ implies that there is a real vector (θ, η, π) such that

$$\theta \cdot A + \eta \cdot B + \pi \cdot E = 0 \text{ and } \eta \geq 0, \pi > 0.$$

Recall that $B_4 = 0$ and $E_4 = 1$, so we obtain that $\theta \cdot A_4 + \pi = 0$.

Let $(q^k)_{k=1}^K$ be vectors of prices such that the dataset $(x^k, q^k)_{k=1}^K$ satisfies SAR-EDU and $\log q_t^k \in \mathbf{Q}$ for all k and s. (Such $(q^k)_{k=1}^K$ exists by Lemma 6.) Construct matrices A', B', and E' from this dataset in the same way as A, B, and E is constructed in the proof of Lemma 5. Note that only the prices are different in (x^k, q^k) compared to (x^k, p^k). So $E' = E$, $B' = B$ and $A'_i = A_i$ for $i = 1, 2, 3$. Since only prices q^k are different in this dataset, only A'_4 may be different from A_4.

By Lemma 6, we can choose prices q^k such that $|\theta \cdot A'_4 - \theta \cdot A_4| < \pi/2$. We have shown that $\theta \cdot A_4 = -\pi$, so the choice of prices q^k guarantees that $\theta \cdot A'_4 < 0$. Let $\pi' = -\theta \cdot A'_4 > 0$.

Note that $\theta \cdot A'_i + \eta \cdot B'_i + \pi' E_i = 0$ for $i = 1, 2, 3$, as (θ, η, π) solves system $S2$ for matrices A, B and E, and $A'_i = A_i$, $B'_i = B_i$ and $E_i = 0$ for $i = 1, 2, 3$. Finally, $B_4 = 0$ so

$$\theta \cdot A'_4 + \eta \cdot B'_4 + \pi' E_4 = \theta \cdot A'_4 + \pi' = 0.$$

We also have that $\eta \geq 0$ and $\pi' > 0$. Therefore θ, η, and π' constitute a solution $S2$ for matrices A', B', and E'.

Lemma 3 then implies that there is no solution to $S1$ for matrices A', B', and E'. So there is no solution to the system comprised by (8), (9), and (10) in the proof of Lemma 5. However, this contradicts Lemma 5 because the data (x^k, q^k) satisfies SAR-EDU and $\log q_t^k \in \mathbf{Q}$ for all $k = 1, \ldots K$ and $t = 1, \ldots, T$.

7. Proof of Theorem 2

The proofs for QHD and PQHD are similar, so we give a detailed proof for PQHD and then explain how the proof for QHD is different.

Lemma 8. *Let $(x^k, p^k)_{k=1}^K$ be a dataset. The following statements are equivalent:*

1. $(x^k, p^k)_{k=1}^K$ *is PQHD rational.*

2. *There are strictly positive numbers* v_t^k, λ^k, $\beta \leq 1$, *and* $\delta \in (0,1]$, *for* $t = 0, \ldots, T$ *and* $k = 1, \ldots, K$, *such that*

$$v_t^k = \lambda^k p_t^k \text{ if } t = 0,\ \beta\delta^t v_t^k = \lambda^k p_t^k \text{ if } t > 0,\ \text{and } x_t^k > x_{t'}^{k'} \Longrightarrow v_t^k \leq v_{t'}^{k'}.$$

The proof of Lemma 8 is very similar to the proof of Lemma 1 and omitted.

7.1. Necessity

Lemma 9. *If a dataset* $(x^k, p^k)_{k=1}^K$ *is PQHD rational, then it satisfies SAR-PQHD.*

Proof. Let $(x^k, p^k)_{k=1}^K$ be PQHD rational, and let $\beta \leq 1$, $\delta \in (0,1]$, and $u : \mathbf{R}_+ \to \mathbf{R}$ be as in the definition of PQHD rational. By Lemma 8, there exists a strictly positive solution v_t^k, λ^k, β, δ to the system in Statement (2) of Lemma 8 with $v_t^k \in \partial u(x_t^k)$ when $x_t^k > 0$, and $v_t^k \geq \underline{w} \in \partial u(x_t^k)$ when $x_t^k = 0$. Moreover, $v_t^k = \lambda^k p_t^k / D(t)$, where $D(t) = 1$ if $t = 0$ and $D(t) = \beta\delta^t$ if $t > 0$.

Let $(x_{t_i}^{k_i}, x_{t'_i}^{k'_i})_{i=1}^n$ be a sequence satisfying the four conditions in SAR-PQHD. Then $x_{t_i}^{k_i} > x_{t'_i}^{k'_i}$. Suppose that $x_{t'_i}^{k'_i} > 0$. Then, $v_{t_i}^{k_i} \in \partial u(x_{t_i}^{k_i})$ and $v_{t'_i}^{k'_i} \in \partial u(x_{t'_i}^{k'_i})$. By the concavity of u, it follows that $v_{t_i}^{k_i} \leq v_{t'_i}^{k'_i}$.Similarly, if $x_{t'_i}^{k'_i} = 0$, then $v_{t_i}^{k_i} \in \partial u(x_{t_i}^{k_i})$ and $v_{t'_i}^{k'_i} \geq \underline{w} \in \partial u(x_{t'_i}^{k'_i})$, so that $v_{t_i}^{k_i} \leq v_{t'_i}^{k'_i}$. Therefore,

$$1 \geq \prod_{i=1}^n \frac{\lambda^{k_i} D(t'_i) p_{t_i}^{k_i}}{\lambda^{k'_i} D(t_i) p_{t'_i}^{k'_i}} = \prod_{i=1}^n \frac{D(t'_i) p_{t_i}^{k_i}}{D(t_i) p_{t'_i}^{k'_i}} = \frac{\beta^{\#\{i:t'_i>0\} - \#\{i:t_i>0\}}}{\delta^{(\sum t_i - \sum t'_i)}} \prod_{i=1}^n \frac{p_{t_i}^{k_i}}{p_{t'_i}^{k'_i}} \geq \prod_{i=1}^n \frac{p_{t_i}^{k_i}}{p_{t'_i}^{k'_i}},$$

where the first equality holds by (4) of SAR-PQHD; and the numbers λ^k appear the same number of times in the denominator as in the numerator of this product. Moreover, the last inequality holds by (2) and (3) of SAR-PQHD. □

7.2. Sufficiency

Lemma 10. *Let data* $(x^k, p^k)_{k=1}^k$ *satisfy SAR-PQHD. Suppose that* $\log(p_t^k) \in \mathbf{Q}$ *for all* k *and* t. *Then there are numbers* v_t^k, λ^k, β, δ, *for* $t \in T$ *and* $k \in K$ *satisfying* (2) *in Lemma 8.*

Lemma 11. *Let data* $(x^k, p^k)_{k=1}^k$ *satisfy SAR-PQHD. Then for all positive numbers* $\overline{\varepsilon}$, *there exists* $q_t^k \in [p_t^k - \overline{\varepsilon}, p_t^k]$ *for all* $t \in T$ *and* $k \in K$ *such that* $\log q_t^k \in \mathbf{Q}$ *and the dataset* $(x^k, q^k)_{k=1}^k$ *satisfy SAR-PQHD.*

Lemma 12. *Let data* $(x^k, p^k)_{k=1}^k$ *satisfy SAR-PQHD. Then there are numbers* v_t^k, λ^k, β, δ, *for* $t \in T$ *and* $k \in K$ *satisfying* (2) *in Lemma 8.*

Lemma 11 and 12 hold as in the proof for Theorem 1.

7.3. Proof of Lemma 10

We linearize the equation in System (2) of Lemma 8. The result is:

$$\log v(x_t^k) - \log \lambda^k - \log p_t^k = 0 \text{ if } t = 0, \tag{11}$$

$$\log v(x_t^k) + \log \beta + t \log \delta - \log \lambda^k - \log p_t^k = 0 \text{ if } t > 0, \tag{12}$$

$$x > x' \Longrightarrow \log v(x') \geq \log v(x), \tag{13}$$

$$\log \beta \geq 0, \tag{14}$$

$$\log \delta \leq 0. \tag{15}$$

In the system comprised by (11), (12) (13), (14) and (15), the unknowns are the real numbers $\log \beta$, $\log \delta$, $\log \lambda^k$, and $\log v_t^k$ for all $k = 1, \ldots, K$ and $t = 1, \ldots, T$.

First, we are going to write the system of inequalities from (11) to (15) in matrix form.

We shall define a matrix A such that there are positive numbers v_t^k, λ^k, β, δ the logs of which satisfy Equations (11) and (12) if and only if there is a solution $u \in \mathbf{R}^{K\times(T+1)+2+K+1}$ to the system of equations

$$A \cdot u = 0,$$

and for which the last component of u is strictly positive.

Let A be a matrix with $K \times (T+1)$ rows and $K \times (T+1) + 2 + K + 1$ columns, defined as follows: We have one row for every pair (k, t); one column for every pair (k, t); two columns for each k; and two additional columns. Organize the columns so that we first have the $K \times (T + 1)$ columns for the pairs (k, t); then two columns, which we shall refer to as the β-column and δ-column, respectively; then K columns (one for each k); and finally one last column. In the row corresponding to (k, t) the matrix has zeroes everywhere with the following exceptions: it has a 1 in the column for (k, t); it has a 1 if $t > 0$ and it has a 0 if $t = 0$ in the β-column; it has a t in the δ-column; it has a -1 in the column for k; and $-\log p_t^k$ in the very last column.

Thus, matrix A looks as follows:

$$
\begin{array}{c|cccccc|c|c|ccccc|c|}
 & (1,1) & \cdots & (k,t) & (k,t') & \cdots & (K,T) & \beta & \delta & 1 & \cdots & k & \cdots & K & p \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\
(k,t=0) & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & t & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\
(k,t'>0) & 0 & \cdots & 0 & 1 & \cdots & 0 & 1 & t' & 0 & \cdots & -1 & \cdots & 0 & -\log p_{t'}^k \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots
\end{array}
$$

Consider the system $A \cdot u = 0$. If there are numbers solving Equations (11) and (12), then these define a solution $u \in \mathbf{R}^{K\times(T+1)+2+K+1}$ for which the last component is 1. If, on the other hand, there is a solution $u \in \mathbf{R}^{K\times(T+1)+2+K+1}$ to the system $A\cdot u = 0$ in which the last component is strictly positive, then by dividing through by the last component of u we obtain numbers that solve Equation (11) and (12).

In second place, we write the system of inequalities (13), (14) and (15) in matrix form. Let B be a matrix with $K \times (T + 1) + 2 + K + 1$ columns. Define B as follows: One row for every pair (k, t) and (k', t') with $x_t^k > x_{t'}^{k'}$; in the row corresponding to (k, t) and (k', t') we have zeroes everywhere with the exception of a -1 in the column for (k, t) and a 1 in the column for (k', t') . Finally, we have last two rows, where we have zero everywhere with one exception. In the first row, we have a -1 at $(K \times (T + 1) + 1)$-th column; in the second row, we have a -1 at $(K \times (T + 1) + 2)$-th column. We shall refer to the first last row as the β-row, which captures (14). We also shall refer to the second row as the δ-row, which captures (15). For (general) QHD, we do not have a β-row.

In third place, we have a matrix E that captures the requirement that the last component of a solution be strictly positive. The matrix E has a single row and $K\times(T+1)+2+K+1$ columns. It has zeroes everywhere except for 1 in the last column.

To sum up, there is a solution to system (11), (12), (13), (14), and (15) if and only if there is a vector $u \in \mathbf{R}^{K\times(T+1)+2+K+1}$ that solves the system of equations and linear inequalities $(S1): \ A \cdot u = 0, \ B \cdot u \geq 0, \ E \cdot u \gg 0$. The argument now follow along the lines of the proof of Theorem 1. Suppose that there is no solution u and let (θ, η, π) be an integer vector solving system $(S2): \ \theta \cdot A + \eta \cdot B + \pi \cdot E = 0, \ \eta \geq 0, \ \pi > 0$.

The following has the same proof as Claim 1.

Claim 9. *(i)* $\theta \cdot A_1 + \eta \cdot B_1 = 0$*; (ii)* $\theta \cdot A_2 + \eta \cdot B_2 = 0$*; (iii)* $\theta \cdot A_3 + \eta \cdot B_3 = 0$*; (iv)* $\theta \cdot A_4 = 0$*; and (v)* $\theta \cdot A_5 + \pi \cdot E_5 = 0$.

We transform the matrices A and B based on the values of θ and η, as we did in the proof of Theorem 1. Lets define a matrix A^* from A and B^* from B, as we did in the proof of Theorem 1. We can prove the same claims (i.e., Claim 2,3,4,5, and 6) as in the proof of Theorem 1. The proofs are the same and omitted. In particular, we can show that there exists a sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ that satisfies (1) in SAR-PQHD. We shall use the sequence of pairs $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ as our candidate violation of SAR-PQHD.

Claim 10. *The sequence $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ satisfies (2), (3), and (4) in SAR-PQHD.*

Proof. We first establish (2). Note that A^*_3 is a vector, and in row r the entry of A^*_3 is as follows. There must be a (k, t) of which r is a copy. Then the component at row r of A^*_3 is t if r is original and $-t$ if r is converted. Now, when r appears as original there is some i for which $t = t_i$, when r appears as converted there is some i for which $t = t'_i$. So for each r there is i such that $(A^*_3)_r$ is either t_i or $-t'_i$.

By Claim 9 (iii), $\theta \cdot A_3 + \eta \cdot B_3 = 0$. Recall that $\theta \cdot A_3$ equals the sum of the rows of A^*_3. Moreover, B_3 is a vector that has zeroes everywhere except a -1 in the δ row (i.e., $K \times (T + 1) + 2$th row). Therefore, the sum of the rows of A^*_3 equals $\eta_{K\times(T+1)+2}$, where $\eta_{K\times(T+1)+2}$ is the $K\times(T+1)+2$th element of η. Since $\eta \geq 0$, therefore, $\sum_{i:t_i>0} t_i - \sum_{i:t'_i>0} t'_i = \eta_{K\times(T+1)+2} \geq 0$, and condition (2) in the axiom is satisfied.

Next, we show (3). By Claim 9 (ii), $\theta \cdot A_2 + \eta \cdot B_2 = 0$. Recall that $\theta \cdot A_2$ equals the sum of the rows of A^*_2. Moreover, B_2 is a vector that has zeroes everywhere except a -1 in the β row (i.e., $K \times (T + 1) + 1$th row). Therefore, the sum of the rows of A^*_2 equals $\eta_{K\times(T+1)+1}$, where $\eta_{K\times(T+1)+1}$ is the $K \times (T + 1) + 1$th element of η. Since $\eta \geq 0$, therefore, $\#\{i : t_i > 0\} - \#\{i : t'_i > 0\} = \eta_{K\times(T+1)+1} \geq 0$, and condition (3) in the axiom is satisfied. (For (general) QHD, B_2 is a zero vector in the β row (i.e., $K \times (T + 1) + 1$th row). Hence, $\#\{i : t_i > 0\} - \#\{i : t'_i > 0\} = 0$, and condition (3) in SAR-QHD is satisfied.)

Now we turn to (4). By Claim 9 (iv), the rows of A^*_4 add up to zero. Therefore, the number of times that k appears in an original row equals the number of times that it appears in a converted row. By Claim 6, then, the number of times k appears as k_i equals the number of times it appears as k'_i. Therefore condition (4) in the axiom is satisfied. □

Finally, we can show that $\prod^{n^*}_{i=1} p^{k_i}_{t_i}/p^{k'_i}_{t'_i} > 1$, which finishes the proof of Lemma 5 as the sequence $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^{n^*}_{i=1}$ would then exhibit a violation of SAR-PQHD. The proof is the same as that of the corresponding lemma in the proof of Theorem 1.

References

Adams, A., L. Cherchye, B. De Rock, and E. Verriest (2014): "Consume Now or Later? Time Inconsistency, Collective Choice and Revealed Preference," *American Economic Review*, 104, 4147–4183.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008): "Eliciting Risk and Time Preferences," *Econometrica*, 76, 583–618.

Andreoni, J., W. T. Harbaugh, and B. J. Gillen (2013a): "The Power of Revealed Preference Tests: Ex-Post Evaluation of Experimental Design," Unpublished manuscript.

Andreoni, J., M. A. Kuhn, and C. Sprenger (2013b): "On Measuring Time Preferences," Unpublished manuscript.

Andreoni, J. and J. Miller (2002): "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 70, 737–753.

Andreoni, J. and C. Sprenger (2012): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102, 3333–3356.

Ashton, L. (2014): "Hunger Games: Does Hunger and Cognitive Fatigue Affect Time Preferences?" Unpublished manuscript.

Augenblick, N., M. Niederle, and C. Sprenger (forthcoming): "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarterly Journal of Economics*.

Barcellos, S. H. and L. Carvalho (2014): "Information about Self-Control and Intertemporal Choices," Unpublished manuscript.

Beatty, T. K. M. and I. A. Crawford (2011): "How Demanding is the Revealed Preference Approach to Demand?" *American Economic Review*, 101, 2782–2795.

Border, K. C. (2013): "Alternative Linear Inequalities," Mimeo, Caltech.

Brocas, I., J. D. Carrillo, and J. Tarrasó (2015): "How Long is a Minute?" Unpublished manuscript.

Bronars, S. G. (1987): "The Power of Nonparametric Tests of Preference Maximization," *Econometrica*, 55, 693–698.

Browning, M. (1989): "A Nonparametric Test of the Life-Cycle Rational Expectations Hypothesis," *International Economic Review*, 30, 979–992.

Carvalho, L., S. Meier, and S. W. Wang (2014): "Poverty and Economic Decision Making: Evidence from Changes in Financial Resources at Payday," Unpublished manuscript.

Carvalho, L., S. Prina, and J. Sydnor (2013): "The Effects of Savings on Risk Attitudes and Intertemporal Choices," Unpublished manuscript.

Chakraborty, A., E. M. Calford, G. Fenig, and Y. Halevy (2014): "Estimating Time Preferences from Convex Budgets: Comment on Normality and Impatience," Unpublished manuscript.

Chambers, C. P. and F. Echenique (2014): "On the Consistency of Data with Bargaining Theories," *Theoretical Economics*, 9, 137–162.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007): "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American economic review*, 97, 1921–1938.

Crawford, I. (2010): "Habits Revealed," *Review of Economic Studies*, 77, 1382–1402.

Crawford, I. and B. De Rock (2014): "Empirical Revealed Preference," *Annual Review of Economics*, 6, 503–524.

Crawford, I. and M. Polisson (2014): "Testing for Intertemporal Nonseparability," *Journal of Mathematical Economics*, 52, 46–49.

Dean, M. and D. Martin (forthcoming): "Measuring Rationality with the Minimum Cost of Revealed Preference Violations," *Review of Economics and Statistics*.

Demuynck, T. and E. Verriest (2013): "I'll Never Forget My First Cigarette: A Revealed Preference Analysis of the "Habits as Durables" Model," *International Economic Review*, 54, 717–738.

Dziewulski, P. (2014): "Revealed Time-Preference," Unpublished manuscript.

Echenique, F. and K. Saito (forthcoming): "Savage in the Market," *Econometrica*.

Fishburn, P. and W. Edwards (1997): "Discount-Neutral Utility Models for Denumerable Time Streams," *Theory and decision*, 43, 139–166.

Fishburn, P. C. and A. Rubinstein (1982): "Time Preference," *International Economic Review*, 23, 677–694.

Fisman, R., S. Kariv, and D. Markovits (2007): "Individual Preferences for Giving," *American Economic Review*, 97, 1858–1876.

Galperti, S. and B. H. Strulovici (2014): "From Anticipations to Present Bias: A Theory of Forward-Looking Preferences," *Available at SSRN 2517742*.

Gilboa, I. (1989): "Expectation and Variation in Multi-Period Decisions," *Econometrica*, 57, 1153–1169.

Giné, X., J. Goldberg, D. Silverman, and D. Yang (2013): "Revising Commitments: Time Preference and Time-Inconsistency in the Field," Unpublished manuscript.

Hayashi, T. (2003): "Quasi-Stationary Cardinal Utility and Present Bias," *Journal of Economic Theory*, 112, 343–352.

Houtman, M. and J. A. H. Maks (1985): "Determining All Maximal Data Subsets Consistent with Revealed Preference," *Kwantitatieve Methoden*, 19, 89–104.

Janssens, W., B. Kramer, and L. Swart (2013): "Time Inconsistent Behavior Under Incomplete Markets," Unpublished manuscript.

Koopmans, T. C. (1960): "Stationary Ordinal Utility and Impatience," *Econometrica*, 28, 287–309.

Kopylov, I. (2010): "Simple Axioms for Countably Additive Subjective Probability," *Journal of Mathematical Economics*, 46, 867–876.

Kuhn, M., P. Kuhn, and M. C. Villeval (2014): "Self Control and Intertemporal Choice: Evidence from Glucose and Depletion Interventions," Unpublished manuscript.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–478.

Liu, E. M., J. Meng, and J. T.-Y. Wang (2014): "Confucianism and Preferences: Evidence from Lab Experiments in Taiwan and China," *Journal of Economic Behavior & Organization*, 104, 106–122.

Lührmann, M., M. Serra-Garcia, and J. Winter (2014): "The Impact of Financial Education on Adolescents' Intertemporal Choices," Unpublished manuscript.

Montiel Olea, J. L. and T. Strzalecki (2014): "Axiomatization and Measurement of Quasi-Hyperbolic Discounting," *Quarterly Journal of Economics*, 129, 1449–1499.

Phelps, E. S. and R. A. Pollak (1968): "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, 35, 185–199.

Rockafellar, R. T. (1997): *Convex Analysis*, Princeton University Press.

Sawada, Y. and Y. Kuroishi (2015): "How Does a Natural Disaster Affect People's Preferencec? An Investigation of a Large Scale Flood in the Philippines using Convex Time Budget Experiments," Unpublished manuscript.

Shaw, W. D., T. Grijalva, J. L. Lusk, and R. Rong (2014): "Does the Convex Time Budget Approach Lead to Evidence of Hyperbolic Discounting when the Time Horizon is Very Long?" Unpublished manuscript.

Stoer, J. and C. Witzgall (1970): *Convexity and Optimization in Finite Dimensions*, Springer-Verlag Berlin.

Weibull, J. W. (1985): "Discounted-Value Representations of Temporal Preferences," *Mathematics of Operations Research*, 10, 244–250.

Online Appendix (not for publication)

A. Proof of Theorem 3

The proof that GTD rational is equivalent to SAR-GTD is identical to the result in Echenique and Saito (forthcoming) with the changes of T to S and $\{D(t)\}_{t\in T}$ to $\{\mu_s\}_{s\in S}$. In the following, we show the proofs for MTD and TSU.

A.1. MTD

The proof that SAR-MTD is equivalent to MTD rationality requires the following modification of the argument in Echenique and Saito (forthcoming).

To see that SAR-MTD is necessary, let $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^n_{i=1}$ be a sequence under the conditions of the axiom. We present the proof under the assumption that u is differentiable, but it is straightforward to use the concavity and the corresponding monotonicity of the superdifferential of u, as we did in the proof of Theorem 1. The first-order condition is $D(t)u'(x^k_t) = \lambda^k p_t$. Then

$$1 \geq \prod_{i=1}^{n} \frac{u'(x^{k_i}_{t_i})}{u'(x^{k'_i}_{t'_i})} = \prod_{i=1}^{n} \frac{\lambda^{k_i} D(t'_i) p^{k_i}_{t_i}}{\lambda^{k'_i} D(t_i) p^{k'_i}_{t'_i}} = \prod_{i=1}^{n} \frac{D(t'_i) p^{k_i}_{t_i}}{D(t_i) p^{k'_i}_{t'_i}} = \prod_{i=1}^{n} \frac{D(t'_i)}{D(t_i)} \prod_{i=1}^{n} \frac{p^{k_i}_{t_i}}{p^{k'_i}_{t'_i}} = \prod_{i=1}^{n} \frac{D(t'_{\pi(i)})}{D(t_i)} \prod_{i=1}^{n} \frac{p^{k_i}_{t_i}}{p^{k'_i}_{t'_i}}.$$

Since $t_i \geq t'_{\pi(i)}$ and D is decreasing it follows that $D(t'_{\pi(i)})/D(t_i) \geq 1$. Therefore we must have that $\prod_{i=1}^{n} p^{k_i}_{t_i}/p^{k'_i}_{t'_i} \leq 1$.

For the proof of sufficiency, consider the setup in the proof of Theorem 1 of Echenique and Saito (forthcoming). Note that the GTD model is the same as the model of subjective expected utility. Let A and B be the matrices as constructed in the proof of Theorem 1 of Echenique and Saito. We need to add rows to B to reflect that $D(t') \geq D(t)$ when $t \geq t'$. To put it precisely, we need an additional row for each pair t, t' such that $t \geq t'$. In the row, we have -1 in the column of t and have 1 in the column of t'. (Remember in the matrix A, we have a column for each $t \in T$, as we do for each $s \in S$ in Echenique and Saito (forthcoming).) In the solution to the dual, we follow the steps of the proof until we construct a balanced sequence $(x^{k_i}_{t_i}, x^{k'_i}_{t'_i})^n_{i=1}$. Such a sequence corresponds to a decomposition of A^* into pairs of rows $(r_i, r'_i)^n_{i=1}$ in which r_i is original and r'_i is converted.

Now consider the column corresponding to t. In the characterization of GTD (and of SEU), the entries in that column are all zero. For MTD, the entries of B are no longer all zero in that column. The sum of the rows of $A^* + B^*$ equal zero. As usual we can eliminate pairs of rows of B such that $1_{t'} - 1_t + 1_t - 1_{t''} = 1_{t'} - 1_{t''}$. So in the matrix B^* all the entries in the column for t will be of the same sign. Let us say that they are all -1.

Recall that each row of A is identified with a pair (k, t). If r is a row say that t *appears* in row r if there is k such that r is the row associated with (k, t). In A^* we may have multiple copies of the same row.

Since the rows of $A^* + B^*$ is zero, the number of times that t appears in an original row minus the number of times that t appears in a converted row equals the number of rows in B^* in which t has a -1. Since we have assumed that there are -1s in the column for t, then there are more original rows in which t appears than converted rows. Let $\pi(i)$ be an arbitrary original row, for each converted row in which t appears. This defines π for all converted rows in which t appears.

There are then original rows in which t appears that are not the image through π of some converted row. For each such row ρ of A^* there is some -1 in B^*, as $A^* + B^* = 0$. Let $\sigma(\rho)$ be the the row of B^* with -1.

The construction is the same for columns t' in which B^* only has 1. Let σ be defined in the same way. This defines π for some rows. For the remaining rows, define π as follows. Let ρ be original, such that t appears in ρ, and t is not in the image of π. There is t' in row $\sigma(\rho)$ with $t' \leq t$ (the row $\sigma(\rho)$ is $1_{t'} - 1_t$). There is a unique converted row ρ' with $\sigma(\rho) = \sigma(\rho')$, a row in which t' appears. So let $\pi(\rho) = \rho'$. This defines π for all rows.

A.2. TSU

The proof that SAR-TSU is equivalent to TSU rationality is similar to the proof of Theorem 1. In the following, we explain the differences.

Lemma 13. *Let $(x^k, p^k)_{k=1}^K$ be a dataset. The following statements are equivalent:*

1. *$(x^k, p^k)_{k=1}^K$ is TSU rational.*
2. *There are strictly positive numbers v_t^k and λ^k for $t = 0, \ldots, T$ and $k = 1, \ldots, K$, such that*

$$v_t^k = \lambda^k p_t^k \text{ and } x_t^k > x_t^{k'} \Longrightarrow v_t^k \leq v_t^{k'}.$$

The proof of Lemma 13 is very similar to the proof of Lemma 1 and omitted.

To see that SAR-TSU is necessary, let $(x_{t_i}^{k_i}, x_{t_i'}^{k_i'})_{i=1}^n$ be a sequence under the conditions of the axiom. We present the proof under the assumption that u_t is differentiable, but it is straightforward to use the concavity and the corresponding monotonicity of the superdifferential of u_t, as we did in the proof of Theorem 1. The first-order condition is $u_t'(x_t^k) = \lambda^k p_t$. Since $t_i = t_i'$ for each i, we obtain

$$1 \geq \prod_{i=1}^{n} \frac{u'_{t_i}(x_{t_i}^{k_i})}{u'_{t_i}(x_{t_i}^{k_i'})} = \prod_{i=1}^{n} \frac{\lambda^{k_i} p_{t_i}^{k_i}}{\lambda^{k_i'} p_{t_i}^{k_i'}} = \prod_{i=1}^{n} \frac{\lambda^{k_i}}{\lambda^{k_i'}} \prod_{i=1}^{n} \frac{p_{t_i}^{k_i}}{p_{t_i}^{k_i'}} = \prod_{i=1}^{n} \frac{p_{t_i}^{k_i}}{p_{t_i}^{k_i'}},$$

where the last equality holds because each k appears as k_i' the same number of times it appears as k_i.

In the following, we prove the sufficiency. The outline of the proof is the same as in the proof of Theorem 1.

Lemma 14. *Let data $(x^k, p^k)_{k=1}^k$ satisfy SAR-TSU. Suppose that $\log(p_t^k) \in \mathbf{Q}$ for all k and t. Then there are numbers v_t^k, λ^k, β, δ, for $t \in T$ and $k \in K$ satisfying* (2) *in Lemma 13.*

Lemma 15. *Let data $(x^k, p^k)_{k=1}^k$ satisfy SAR-TSU. Then for all positive numbers $\overline{\varepsilon}$, there exists $q_t^k \in [p_t^k - \overline{\varepsilon}, p_t^k]$ for all $t \in T$ and $k \in K$ such that $\log q_t^k \in \mathbf{Q}$ and the dataset $(x^k, q^k)_{k=1}^k$ satisfy SAR-TSU.*

Lemma 16. *Let data $(x^k, p^k)_{k=1}^k$ satisfy SAR-TSU. Then there are numbers v_t^k and λ^k for all $t \in T$ and $k \in K$ satisfying* (2) *in Lemma 13.*

Lemma 15 and 16 hold as in the proof for Theorem 1.

A.3. Proof of Lemma 14

We linearize the equation in System (2) of Lemma 13. The result is:

$$\log v_t(x_t^k) - \log \lambda^k - \log p_t^k = 0, \tag{16}$$

$$x_t^k > x_t^{k'} \Longrightarrow \log v_t(x_t^k) \leq \log v_t(x_t^{k'}). \tag{17}$$

In the system comprised by (16) and (17), the unknowns are the real numbers λ^k and $\log v_t^k$ for all $k = 1, \ldots, K$ and $t = 1, \ldots, T$.

We shall define a matrix A such that there are positive numbers v_t^k and λ^k, the logs of which satisfy Equation (16) if and only if there is a solution $u \in \mathbf{R}^{K \times (T+1)+K+1}$ to the system of equations

$$A \cdot u = 0,$$

and for which the last component of u is strictly positive.

Let A be a matrix with $K \times (T+1)$ rows and $K \times (T+1) + K + 1$ columns. The matrix A is similar to the matrix A defined in the proof of Theorem 1. Only the difference here is that we no longer have the δ-column. Thus, matrix A looks as follows:

$$
\begin{array}{c}
\\
\vdots \\
{\scriptstyle (k,t)} \\
\vdots
\end{array}
\begin{array}{c}
\begin{array}{ccccc|ccccc|c}
{\scriptstyle (1,0)} & {\scriptstyle \cdots} & {\scriptstyle (k,t)} & {\scriptstyle \cdots} & {\scriptstyle (K,T)} & {\scriptstyle 1} & {\scriptstyle \cdots} & {\scriptstyle k} & {\scriptstyle \cdots} & {\scriptstyle K} & {\scriptstyle p}
\end{array} \\
\left[
\begin{array}{ccccc|ccccc|c}
\vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & 1 & \cdots & 0 & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\
\vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots
\end{array}
\right]
\end{array}
$$

Consider the system $A \cdot u = 0$. If there are numbers solving Equation (16), then these define a solution $u \in \mathbf{R}^{K \times (T+1)+K+1}$ for which the last component is 1. If, on the other hand, there is a solution $u \in \mathbf{R}^{K \times (T+1)+K+1}$ to the system $A \cdot u = 0$ in which the last component is strictly positive, then by dividing through by the last component of u we obtain numbers that solve Equation (16).

In second place, we write the system of inequality (17) in matrix form. Let B be a matrix with $K \times (T+1) + K + 1$ columns. Define B as follows: One row for every pair (k,t) and (k',t) with $x_t^k > x_t^{k'}$; in the row corresponding to (k,t) and (k',t) we have zeroes everywhere with the exception of a -1 in the column for (k,t) and a 1 in the column for (k',t).

In third place, we have a matrix E that captures the requirement that the last component of a solution be strictly positive. The matrix E has a single row and $K \times (T+1) + K + 1$ columns. It has zeroes everywhere except for 1 in the last column.

To sum up, there is a solution to system (16) and (17) if and only if there is a vector $u \in \mathbf{R}^{K \times (T+1)+K+1}$ that solves the system of equations and linear inequalities

$$(S1): \ A \cdot u = 0, \ B \cdot u \geq 0, \ E \cdot u \gg 0.$$

The entries of A, B, and E are integer numbers, with the exception of the last column of A. Under the hypothesis of the lemma we are proving, the last column consists of rational numbers.

By Lemma 4, then, there is such a solution u to $S1$ if and only if there is no vector (θ, η, π) that solves the system of equations and linear inequalities

$$(S2): \ \theta \cdot A + \eta \cdot B + \pi \cdot E = 0, \ \eta \geq 0, \ \pi > 0.$$

In the following, we shall prove that the non-existence of a solution u implies that the data must violate SAR-TSU. Suppose then that there is no solution u and let (θ, η, π) be a rational vector as above, solving system $S2$.

By multiplying (θ, η, π) by any positive integer we obtain new vectors that solve $S2$, so we can take (θ, η, π) to be integer vectors.

For convenience, we transform the matrices A and B using θ and η. We now transform the matrices A and B based on the values of θ and η, as we did in the proof of Theorem 1. Lets define a matrix A^* from A and B^* from B, as we did in the proof of Theorem 1. We can prove the same claims (i.e., Claim 2,3,4,5, and 6) as in the proof of Theorem 1. The proofs are the same and omitted. In particular, we can show that there exists a sequence of pairs $(x_{t_i}^{k_i}, x_{t'_i}^{k'_i})_{i=1}^{n^*}$ that satisfies (1) in SAR-TSU. Moreover, by the definition of B matrix, we have $t_i = t'_i$ because in matrix B we have $z >^i z'$ if there exist $t \in T$ and $k, k' \in T$ such that there exist $x_t^k = z$ and $x_t^{k'} = z'$. Moreover, as in Claim 10, we can show that in the sequence $(x_{t_i}^{k_i}, x_{t'_i}^{k'_i})_{i=1}^{n^*}$, each k appears k_i the same number of times it appears as k'_i. Finally, we can show that $\prod_{i=1}^{n^*} p_{t_i}^{k_i}/p_{t'_i}^{k'_i} > 1$, which finishes the proof of Lemma 14 as the sequence $(x_{t_i}^{k_i}, x_{t'_i}^{k'_i})_{i=1}^{n^*}$ would then exhibit a violation of SAR-TSU. The proof is the same as in the proof of Theorem 1 and omitted.

B. Implementing Revealed Preference Tests

This section presents a method to implement the revealed preference tests for time discounting models using Matlab R2014b (MathWorks). We use Andreoni and Sprenger's (2012) experimental choice data as the model case, but our method is applicable to other empirical/experimental data sets.

Dataset. Subjects in the Andreoni and Sprenger's (2012) experiment completed 45 intertemporal decisions with varying starting dates τ, delay lengths d, and gross interest rates $a_{\tau+d}/a_\tau$ and, in particular, they complete 5 decision problems for each pair of (τ, d). See Figure B.1 for an illustration of budgets. For each subject, decision in every trial is characterized by a tuple $(\tau, d, a_\tau, a_{\tau+d}, c_\tau)$ where c_τ is the number of tokens allocated to sooner payment.

The following figure illustrates the budgets faced by the subjects in AS's experiment, fixing one time frame at (τ, d).

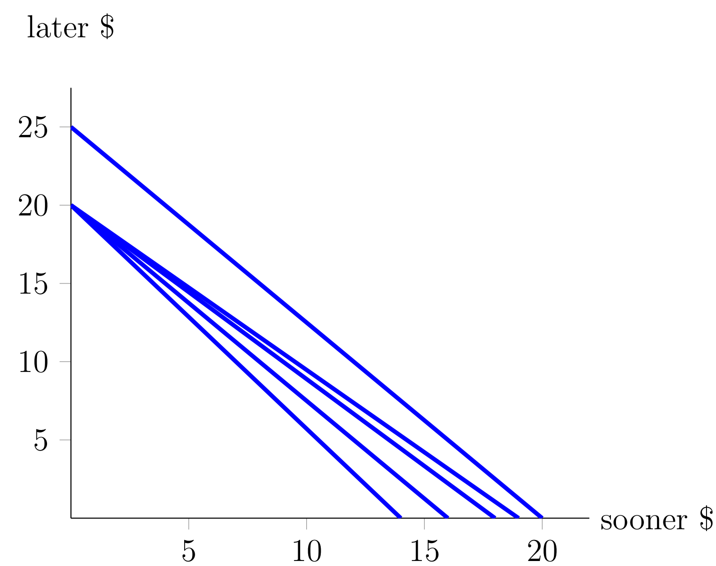


Figure B.1: An illustration of the CTB design in Andreoni and Sprenger (2012). Budget sets are represented in blue lines, fixing one time frame at $(\tau, d) = (0, 35)$.

In order to rewrite our data in price-consumption format as in the theory, we set prices $p_\tau = 1 + r = a_{\tau+d}/a_\tau$ and $p_{\tau+d} = 1$ (normalization), and define consumptions $x_\tau = c_\tau \cdot a_\tau$ and $x_{\tau+d} = (100 - c_\tau) \cdot a_{\tau+d}$. This gives us a dataset $(x^k, p^k)_{k=1}^{45}$.

As we explained in the main body of the paper, we implicitly set prices of consumption in periods that were not offered to a subject as very high in order to ensure that consumption is zero. The idea is as follows. Think of EDU for concreteness. We use first order conditions, so that we are looking for a rationalizing u and δ such that $\delta^t u'(x_t^k) = \lambda^k p_t^k$ if $x_t^k > 0$ and $\delta^t u'(x_t^k) \leq \lambda^k p_t^k$ if $x_t^k = 0$. In setting up such a system of equations we can ignore the t that was not offered to the agents in trial k. Then whatever u we construct will have a finite derivative $u'(0)$ at zero. Hence we can set p_t^k to be high enough so that the agent finds it optimal to consume $x_t^k = 0$. By this argument is it clear that one can ignore the (zero) consumption in the periods that were not offered in trial k, we think of consumption in those periods as prohibitively expensive. This is of course consistent with the fact that AS did not offer subjects any consumption in those periods; consumption in those periods is infeasible. The set of time periods we are looking at is thus $T = \{0, 7, 35, 42, 70, 77, 98, 105, 133\}$.

We are able to check whether a given dataset is consistent with TSU, GTD, MTD, QHD, PQHD, or EDU, by solving the corresponding linear programming problem. The construction of linear programming problems closely follows the argument in the proofs of Theorem 1, 2, and 3. In particular, the key to this procedure is to set up a system of linear inequalities of the form:

$$S : \begin{cases} A \cdot u = 0 \\ B \cdot u \geq 0 \\ E \cdot u > 0 \end{cases}$$

which, in the case of EDU for example, is a matrix form of the linearized system:

$$\begin{aligned} \log v(x_t^k) + t \log \delta - \log \lambda^k - \log p_t^k = 0, \\ x > x' \Longrightarrow \log v(x') \geq \log v(x), \\ \log \delta \leq 0. \end{aligned}$$

A system of linear inequalities. We now construct three key ingredients of the system, matrices A, B, and E, starting from those necessary for testing EDU. The first matrix A looks as follows:

$$A = \begin{array}{c} \\ \vdots \\ (k,t) \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc|c|ccccc|c} (1{,}0) & \cdots & (k{,}t) & \cdots & (45{,}133) & \delta & 1 & \cdots & k & \cdots & 45 & p \end{array} \\ \left[\begin{array}{ccccc|c|ccccc|c} \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & t & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\ \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \end{array}\right] \end{array}.$$

Since we can ignore the t that was not offered to the agents in trial k, the matrix has $45 \times 2 = 90$ rows and $45 \times 2 + 1 + 45 + 1 = 137$ columns. In the row corresponding to (k, t)

the matrix has zeroes everywhere with the following exceptions: it has a 1 in the column for (k,t); it has a t in the δ column; it has a -1 in the column for k; and $-\log p_t^k$ in the very last column. This finalizes the construction of A.

Next, we construct matrix B that has 137 columns and there is one row for every pair (k,t) and (k',t') with $x_t^k > x_{t'}^{k'}$. In the row corresponding to (k,t) and (k',t') we have zeroes everywhere with the exception of a -1 in the column for (k,t) and a 1 in the column for (k',t'). Finally, in the last row, we have zero everywhere with the exception of a -1 at 91st column. We shall refer to this last row as the δ-row.

Finally, we prepare a matrix that captures the requirement that the last component of a solution be strictly positive. The matrix E has a single row and 137 columns. It has zeroes everywhere except for 1 in the last column.

Constructing matrices for other tests. In order to test models other than EDU, we need to modify matrices A, B, and E appropriately.

For the QHD test, we insert another column capturing the present/future bias parameter β, which we shall refer to the β-column. Hence, three matrices A, B, and E have $45 \times 2 + 1 + 1 + 45 + 1 = 138$ columns. In the row corresponding to (k,t) of the matrix A, the β-column has a 1 if $t > 0$ and a 0 if $t = 0$, indicating "now" or "future".

$$A = \begin{array}{c} \\ \vdots \\ {\scriptstyle (k,t=0)} \\ {\scriptstyle (k,t'>0)} \\ \vdots \end{array} \left[\begin{array}{cccccc|c|c|ccccc|c} {\scriptstyle (1,1)} & \cdots & {\scriptstyle (k,t)} & {\scriptstyle (k,t')} & \cdots & {\scriptstyle (45,133)} & {\scriptstyle \beta} & {\scriptstyle \delta} & {\scriptstyle 1} & \cdots & {\scriptstyle k} & \cdots & {\scriptstyle K} & {\scriptstyle p} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & t & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\ 0 & \cdots & 0 & 1 & \cdots & 0 & 1 & t' & 0 & \cdots & -1 & \cdots & 0 & -\log p_{t'}^k \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \end{array} \right].$$

The construction of matrix B for testing general QHD is the same as above (although the size is now different). For the PQHD test, we add β-row which has 0 everywhere except -1 in the β-column to capture $\beta \leq 1$.

For the MTD and GTD tests, we have 9 columns capturing time-varying discount factors $D(t)$'s.

$$A = \begin{array}{c} \\ \vdots \\ {\scriptstyle (k,t)} \\ \vdots \end{array} \left[\begin{array}{ccccc|ccc|ccccc|c} {\scriptstyle (1,0)} & \cdots & {\scriptstyle \widetilde{x}_\ell} & \cdots & {\scriptstyle (45,133)} & \cdots & {\scriptstyle D(t)} & \cdots & {\scriptstyle 1} & \cdots & {\scriptstyle k} & \cdots & {\scriptstyle 45} & {\scriptstyle p} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \end{array} \right].$$

In the matrix B, we add rows

$$\begin{array}{c} \\ \\ \\ \end{array}\begin{array}{c} \begin{array}{ccccc|cccc|ccccc|c} {\scriptstyle (1,0)} & {\scriptstyle \cdots} & {\scriptstyle (k,t)} & {\scriptstyle \cdots} & {\scriptstyle (45,133)} & {\scriptstyle \cdots} & {\scriptstyle D(t)} & {\scriptstyle D(t+1)} & {\scriptstyle \cdots} & {\scriptstyle 1} & {\scriptstyle \cdots} & {\scriptstyle k} & {\scriptstyle \cdots} & {\scriptstyle 45} & {\scriptstyle p} \end{array} \\ \left[\begin{array}{ccccc|cccc|ccccc|c} \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & -1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \end{array}\right] \end{array}$$

in testing MTD to impose the monotonicity restriction on $D(t)$'s.

The matrix A for testing TSU is similar to that appears in testing EDU. The difference is that we no longer have the δ-column.

$$\begin{array}{c} \\ \vdots \\ {\scriptstyle (k,t)} \\ \vdots \end{array}\begin{array}{c} \begin{array}{ccccc|ccccc|c} {\scriptstyle (1,0)} & {\scriptstyle \cdots} & {\scriptstyle (k,t)} & {\scriptstyle \cdots} & {\scriptstyle (K,T)} & {\scriptstyle 1} & {\scriptstyle \cdots} & {\scriptstyle k} & {\scriptstyle \cdots} & {\scriptstyle K} & {\scriptstyle p} \end{array} \\ \left[\begin{array}{ccccc|ccccc|c} \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & 0 & \cdots & -1 & \cdots & 0 & -\log p_t^k \\ \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \end{array}\right] \end{array}$$

Next, we construct B as follows: One row for every pair (k,t) and (k',t) with $x_t^k > x_t^{k'}$; in the row corresponding to (k,t) and (k',t) we have zeroes everywhere with the exception of a -1 in the column for (k,t) and a 1 in the column for (k',t).

Solve the system. Our task is to check if there is a vector u that solves the following system of linear inequalities corresponding to a model M

$$S_M : \begin{cases} A \cdot u = 0 \\ B \cdot u \geq 0 \\ E \cdot u > 0 \end{cases} .$$

If there is a solution u to this system, we say that the dataset is M rational.

We use the function `linprog` in the Optimization Toolbox of Matlab to find a solution. More precisely, we translate the systems of linear inequalities S_M into constraints in a linear programming problem and solve

$$LP_M : \begin{cases} \min & z \cdot u \\ \text{s.t.} & A \cdot u = 0 \\ & -B \cdot u \leq 0 \\ & -E \cdot u < 0 \end{cases}$$

where z is a zero vector.

It is not possible, however, to specify strict inequality constraints in `linprog`. As an alternative, we find a solution u that has 1 in the last element, i.e., $u_p = 1$. In other words, we solve a normalized version of the problem,

$$LP'_M : \begin{cases} \min & z \cdot u \\ \text{s.t.} & A \cdot u = 0 \\ & -B \cdot u \leq 0 \\ & u_p = 1 \end{cases}$$

where z is a zero vector as above. Here, the constraint $E \cdot u > 0$ is omitted since it is automatically satisfied by our normalization $u_p = 1$.

If the given dataset is EDU rational, we can recover upper and lower bounds of the daily discount factor consistent with the observed choice data. Remember that we include the δ-row in B. The constraint $B \cdot u \geq 0$ then implies that the 91st element of any solution u^* of LP'_M, called u^*_δ, captures the daily discount factor. To be more precise, we can recover the daily discount factor δ by $\exp(u^*_\delta)$ since we normalize u^*_p to be 1. Hence, a solution (if any) of LP'_M in which the 91st element of z is 1 and 0 elsewhere suggests an lower bound of δ and a solution (if any) of LP'_M in which the 91st element of z is -1 and 0 elsewhere suggests an upper bound of δ. In a similar manner, we can recover bounds of present/future biasedness β.

C. Ground Truth Analysis: Test Performance and Parameter Recovery

We assess the basic performance of our revealed preference tests using simulated choices. As in Andreoni and Sprenger (2012), we assume a decision maker has a utility function (CRRA with quasi-hyperbolic discounting) of the form:

$$U(x_0, \ldots, x_T) = \frac{1}{\alpha}x_0^{\alpha} + \beta \sum_{t \in T \setminus \{0\}} \frac{1}{\alpha}\delta^t x_t^{\alpha}.$$

We simulate synthetic subjects' choice data in Andreoni and Sprenger's (2012) environment (i.e., time frames and budgets are identical to those actual subjects faced in their experiment) under all combinations of parameters $\alpha \in \{0.8, 0.82, \ldots, 1\}$, $\delta \in \{0.95, 0.951, \ldots, 1\}$, and $\beta \in \{0.8, 0.82, \ldots, 1.2\}$, resulting the total of 11,781 such synthetic subjects. We then perform our revealed preference tests, in particular, tests for EDU and QHD rationality, and ask following questions: (i) Do our tests correctly identify EDU or QHD rational datasets?; and (ii) Can our tests recover "true" underlying model parameters?

A few remarks are in order. (1) For some parameter specifications, it is possible that the slope of (linear) indifference curves coincide with those of budget lines. This happens 21 times when $(\alpha, \delta) = (1, 1)$. [20] If the slope of indifference curve coincides with the budget line (i.e., every point on the budget yields the same level of utility), we randomly pick one point from the budget as the optimal choice as a tie-breaking rule. (2) In order to avoid rounding issue in Matlab, we treat numbers less than 10^{-10} to be 0. In other words, if the predicted allocation is sufficiently close to a corner, we treat it as a corner choice. (3) Unlike Andreoni and Sprenger's (2012) original experiment where subjects made choices from "discrete" budget sets by allocating 100 tokens, we allow simulated choices to be at any point on the continuous budget lines. We also prepare another set of simulated choices (with the same set of parameters) which mimic behavior of the Andreoni and Sprenger's (2012) experimental subjects for the purpose of comparison.

Test results. The results are presented in Table C.1. We first look at our baseline simulation in which choices were made from continuous budget sets. Of the 11,781 synthetic subjects, 3,950 (33.5%) passed the EDU test and 11,781 (100.0%) passed the QHD test.

[20] For example, consider the case when $(\alpha, \delta, \beta) = (1, 1, 0.8)$ and $(1, 1, 0.9)$. Since the utility function has the form $x_\tau + \beta x_{\tau+d}$ when $\tau = 0$, indifference curve coincides with budget line when prices are 1.11 or 1.25. Another possibility is in the time frame $(\tau, d) = (7, 70)$, where the price of 1 (tokens allocated to sooner and later payments have the same exchange rate) is offered. In this case, indifference curve coincides with budget line as long as $(\alpha, \delta) = (1, 1)$.

TABLE C.1: Test results using simulated choice data from continuous budgets (top panel) and discrete budgets (bottom panel).

	Parameters			
	$\alpha = 1$	$\alpha < 1$	$\alpha < 1$	
Continuous budget		$\beta = 1$	$\beta \neq 1$	Total
No interior choice	1,050	38	700	1,788
Pass EDU	939	510	2,501	3950
Pass QHD	1,071	510	10,200	11,781
Sample size	1,071	510	10,200	11,781

	Parameters			
	$\alpha = 1$	$\alpha < 1$	$\alpha < 1$	
Discrete budget		$\beta = 1$	$\beta \neq 1$	Total
No interior choice	1,050	252	4,746	6,048
Pass EDU	939	510	6,913	8,362
Pass QHD	1,071	510	8,319	9,900
Sample size	1,071	510	10,200	11,781

We then split the sample into three groups. The first group of subjects have linear utility function ($\alpha = 1$). They made no interior choices (except for the knife edge case described above), and 939 of them passed the EDU test. The second group of subjects have nonlinear utility and no present/future bias ($\alpha < 1$, $\beta = 1$). They all passed the EDU test (and hence QHD test, too), as expected. The third group of subjects have nonlinear utility and present/future bias ($\alpha < 1$, $\beta \neq 1$). We find that 2,501 of them passed the EDU test, even though their underlying preferences were strictly present/future biased.

The bottom panel of Table C.1 presents the results with simulated data when choices are assumed to be on the discrete points on the budget lines. As one can imagine, the number of synthetic subjects who make no interior choices increases and accordingly the pass rate for the EDU test increases from 33.5% to 71.0%. We also find that "perturbations" induced by discretization of budget sets is powerful enough for some of the subjects to become QHD non rational.

Parameter recovery. Next we investigate how precise we can recover underlying preference parameters using our revealed preference tests. Remember that the revealed preference tests boil down to linear programming problems. As we describe in Section B of the Online

Appendix, we can find bounds of daily discount factor δ or present-biasedness β, which can be used to rationalize the observed choice data.

In this exercise we restrict our attention to the case of choices from continuous budgets.

1. We look at subset of synthetic subjects who have non-linear instantaneous utility ($\alpha < 1$), no present/future bias ($\beta = 1$), and pass the EDU test. We exclude synthetic subjects who make no interior allocation from this sample. There are 510 subjects in this category. Of those, 304 have $(0, \bar{\delta}_i]$ for some $\bar{\delta}_i < 1$, 10 have $[\underline{\delta}_i, 1]$ for some $\underline{\delta}_i > 0$, and 196 have $[\underline{\delta}_i, \bar{\delta}_i]$ for some combination of $\bar{\delta}_i < 1$ and $\underline{\delta}_i > 0$. Furthermore, within the last category of subjects, the true underlying discount factors are always covered by the ranges $[\underline{\delta}_i, \bar{\delta}_i]$. See Figure C.1, left panel.

2. We focus on those who have non-linear instantaneous utility ($\alpha < 1$), present/future bias ($\beta \neq 1$), and pass the QHD test. We exclude synthetic subjects who make no interior allocation from this sample. There are 10,200 subjects in this category. Of those, 9,737 have $(0, \bar{\delta}_i]$ for some $\bar{\delta}_i < 1$, 43 have $[\underline{\delta}_i, 1]$ for some $\underline{\delta}_i > 0$, and 420 have $[\underline{\delta}_i, \bar{\delta}_i]$ for some combination of $\bar{\delta}_i < 1$ and $\underline{\delta}_i > 0$. Within the last category of subjects, the true underlying discount factors are covered by the ranges $[\underline{\delta}_i, \bar{\delta}_i]$ in 362 cases (86.2%). See Figure C.1, right panel. Next we turn to present bias. Of total 10,200 in this sample, 700 have $(0, \infty)$ (i.e., any value is possible), 76 have $(0, \bar{\beta}_i]$ for some $\bar{\beta}_i > 0$, 60 have $[\underline{\beta}_i, \infty)$ for some $\underline{\beta}_i > 0$, and 9,364 have $[\underline{\beta}_i, \bar{\beta}_i]$ for some combination of $\bar{\beta}_i, \underline{\beta}_i > 0$. Within the last category of subjects, the true underlying present biases are covered by the ranges $[\underline{\beta}_i, \bar{\beta}_i]$ in 9,354 cases (99.9%). See Figure C.2.

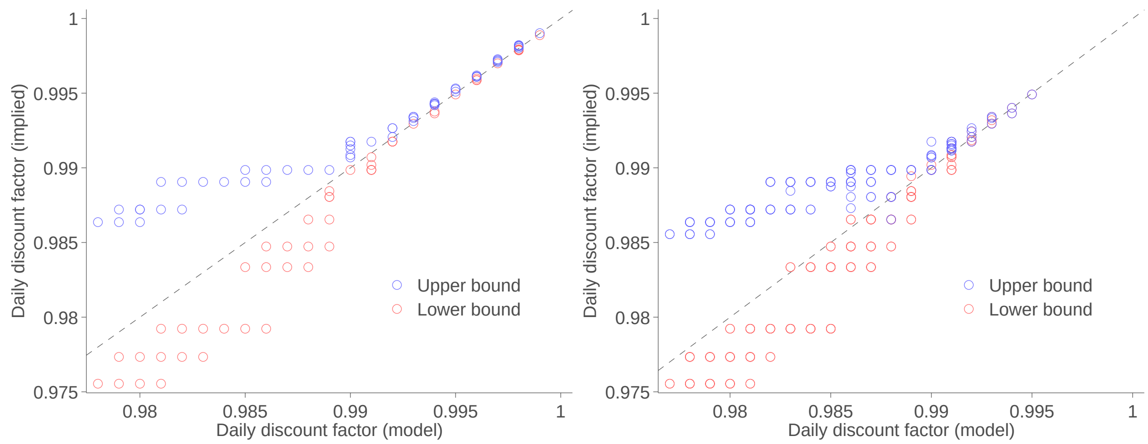


FIGURE C.1: Upper and lower bounds of daily discount factor implied by the revealed preference test. Each synthetic subject has one pair of a blue circle (upper bound) and a red circle (lower bound). (Left) The sample is 196 synthetic subjects who (i) have non-linear utility ($\alpha < 1$) and no present/future bias ($\beta = 1$), (ii) pass the EDU test, and (iii) have recovered range $[\underline{\delta}_i, \bar{\delta}_i]$ with $\underline{\delta}_i > 0$ and $\bar{\delta}_i < 1$. (Right) The sample is 420 synthetic subjects who (i) have non-linear utility ($\alpha < 1$) and present/future bias ($\beta \neq 1$) and (iii) have recovered range $[\underline{\delta}_i, \bar{\delta}_i]$ with $\underline{\delta}_i > 0$ and $\bar{\delta}_i < 1$. The dotted line represents the 45-degree line.

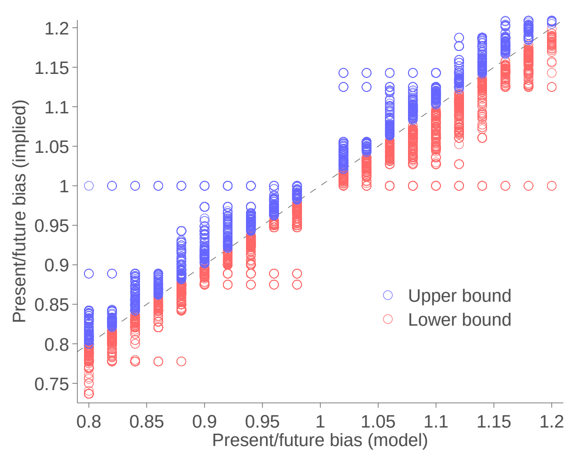


FIGURE C.2: Upper and lower bounds of present/future biadnesss implied by the revealed preference test. Each synthetic subject has one pair of a blue circle (upper bound) and a red circle (lower bound). The sample is 9,364 synthetic subjects who (i) have non-linear utility ($\alpha < 1$) and present/future bias ($\beta \neq 1$), (ii) pass the QHD test, and (iv) have recovered range $[\underline{\beta}_i, \bar{\beta}_i]$ with $\underline{\beta}_i, \bar{\beta}_i > 0$. The dotted line represents the 45-degree line.

D. Additional Results from Empirical Application

In this section we provide additional results supporting the argument in Section 5.2 where we compare AS's parametric estimation of a QHD model and results from our nonparametric revealed preference tests and present our measure of distance from M rationality.

D.1. Comparing EDU Rational and Non-Rational subjects

Remember that AS estimate the per-period discount factor, present bias, and utility curvature assuming a QHD model with CRRA utility over money:

$$U(x_0, \ldots, x_T) = \frac{1}{\alpha}x_0^\alpha + \beta \sum_{t\in T\setminus\{0\}} \delta^t \frac{1}{\alpha}x_t^\alpha.$$

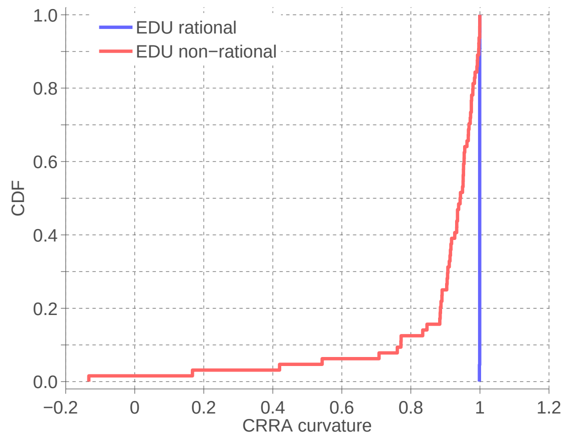
Here we focus on AS's individual level nonlinear least squares (NLS) estimation.

We classify subjects in two groups, those who violate and those who satisfy EDU based on the revealed preference tests. Panels (A)-(C) of Figure D.1 present empirical cumulative distribution functions (CDFs) for the estimated preference parameters in the EDU rational and EDU non-rational groups. Similarly, panels (D)-(F) compare properties of individual's choices (e.g., proportion of interior choices) for the same two groups of subjects.

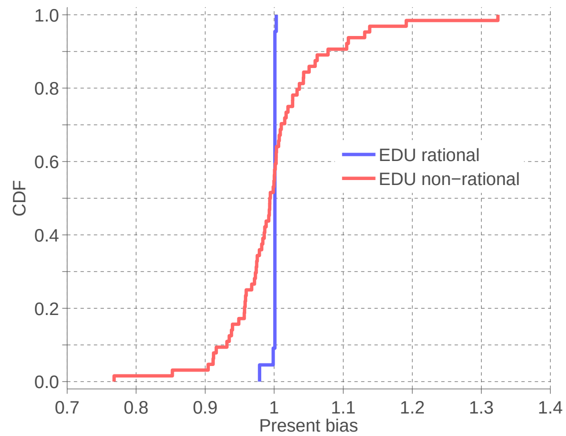
The figure shows how our test is consistent with AS's estimates. Consider panel (B). The CDF for EDU rational subjects concentrates a large mass at $\beta = 1$. The non-EDU group has no such jump in mass at $\beta = 1$, and instead exhibits a substantial fraction of subjects with estimated β different from 1. The CDF for EDU-rational subjects is significantly different from the CDF for EDU non-rational subjects: The null hypothesis of equality-of-distribution is rejected by the two-sample Kolmogorov-Smirnov test ($p < 0.01$).

Figure D.1 panel (B) also shows that subjects who fail our EDU test have estimates of β that differ clearly from 1. An OLS regression of the absolute difference between estimated present bias and 1, $|\widehat{\beta} - 1|$, on a dummy variable for EDU rationality (takes 1 if that subject fails the EDU test) reveals that $\widehat{\beta}$ for EDU non-rational subjects is further away from 1 compared to EDU rational subjects (Table D.1, column 1). Similar result holds for subjects who are not EDU rational but TSU rational and those who are not TSU rational (Table D.1, column 2).

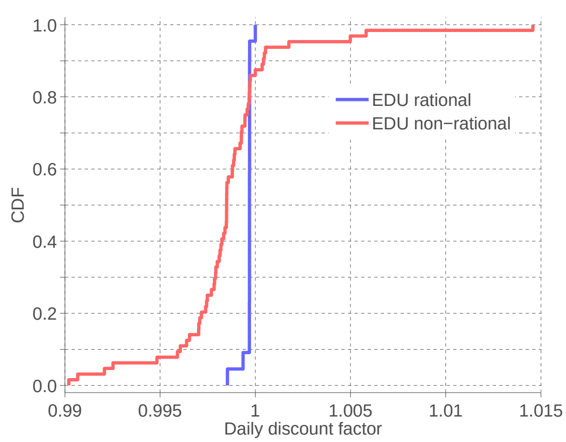
However, $\beta \neq 1$ is not immediately translated into evidence for present or future bias. As we have shown above, most of the subjects who fail the EDU test also fail the QHD test (no additional subject passes the test for PQHD, and most of the subjects who failed EDU even fail MTD). In this sense, the interpretation of estimated β for EDU non-rational subjects



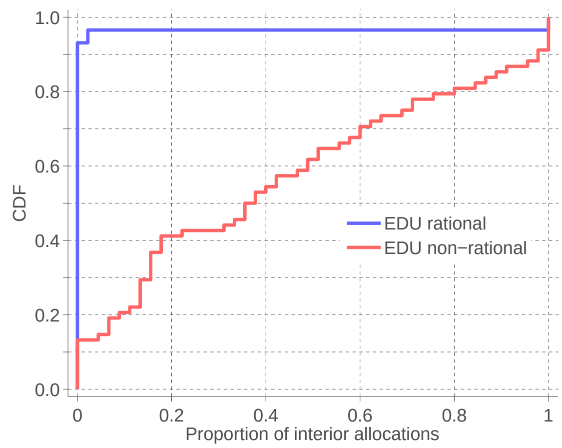
(A) CRRA curvature



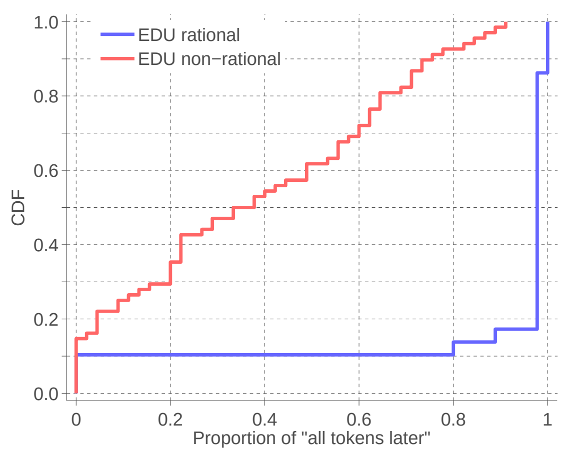
(B) Present bias



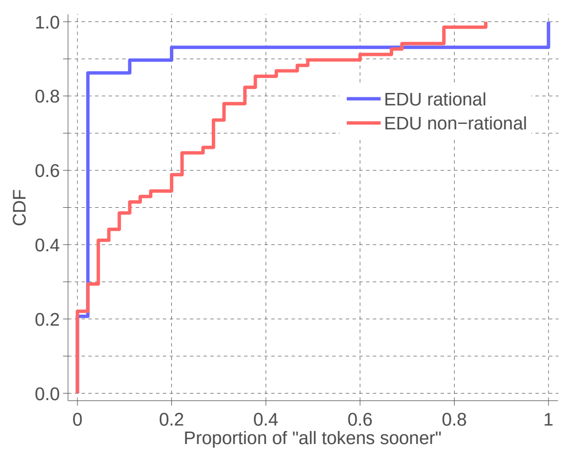
(C) Daily discount factor



(D) Proportion of interior allocations



(E) Proportion of “all later” allocations



(F) Proportion of “all sooner” allocations

FIGURE D.1: Empirical CDFs for preference parameters and properties of choices. Panels (A)-(C) include 86 subjects whose preference parameters are estimable. Panels (D)-(F) include all 97 subjects.

TABLE D.1: OLS regression of $|\widehat{\beta} - 1|$ on rationality dummies.

	(1)	(2)
nonEDU	0.046 ***	
	(0.007)	
$TSU \backslash EDU$		0.055 ***
		(0.019)
nonTSU		0.043 ***
		(0.007)
Constant	0.002 **	0.002 **
	(0.001)	(0.001)
R^2	0.139	0.147
# Obs.	86	86

Notes: *nonEDU* is a dummy for subjects who fail the EDU test, $TSU \backslash EDU$ is a dummy for those who fail the EDU test but pass the TSU test, and *nonTSU* is a dummy for those who pass the TSU test. Robust standard errors are reported in parentheses. Level of significance. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.10$.

in Figure D.1 panel (B) requires some caution. The model is arguably misspecified for such subjects.

One of the advantages of our revealed preference tests is that we can go beyond the class of QHD utility function by weakening the restrictions in the relevant revealed preference axioms.

Consider Figure 2 again. It is interesting to note that the estimated values of $\widehat{\beta}$ for subjects who fail our EDU test are symmetrically distributed around 1.[21] The "average" subject looks, in some sense, as an EDU agent, even though the majority of subjects are not consistent with that model according to our test. It is therefore possible that AS's finding in favor of EDU in their aggregate preference estimation reflects the choice behavior of such an average subject.

[21]We test symmetry using the two-sample Kolmogorov-Smirnov (K-S) test. We first sort estimated $\widehat{\beta}$ in an ascending order, calculate $|\widehat{\beta} - 1|$, and split them into the first half (smaller $\widehat{\beta}$) and the last half (larger $\widehat{\beta}$). We apply K-S test for equality of distribution for those two empirical distributions of $|\widehat{\beta} - 1|$. The null hypothesis of equal distribution is not rejected ($p = 0.132$).

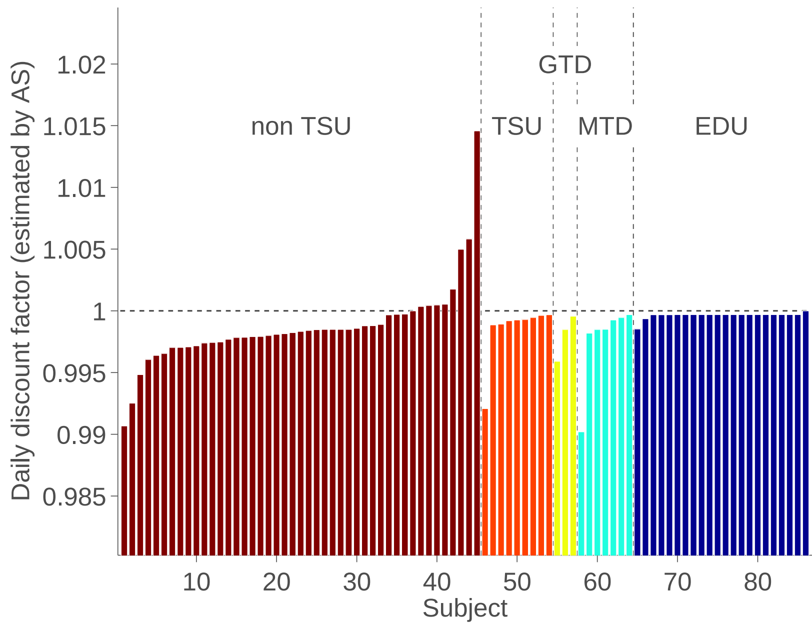


FIGURE D.2: Estimated daily discount factor for each category of subjects.

D.2. Estimated Daily Discount Factors

As in Figure 2 where we show AS's estimated present-bias parameter $\widehat{\beta}$ for each class of rationality, Figure D.2 demonstrates the similar comparison for the case of AS's estimated daily discount factor $\widehat{\delta}$. The subjects who pass the EDU test have estimated $\widehat{\delta}$ very close to 1 (many of them have $\widehat{\delta} = 0.9997$). The subjects who do not pass any of the tests (i.e., non TSU subjects) have estimated $\widehat{\delta}$ which are far from 1 in magnitude compared to the other groups of subjects. Furthermore, those who have $\widehat{\delta} > 1$ are all in this category.

D.3. Parameter Recovery

As we describe in Section B of the Online Appendix, we can find bounds of daily discount factor δ or present-biasedness β, which can rationalize the observed choice data.

Table D.2 lists bounds of discount factor (together with estimated values provided by AS) for 29 EDU rational subjects, and Table D.3 lists bounds of present-biasedness for the same 29 QHD rational subjects.

Table D.2: Recovered bounds for daily discount factor (29 EDU rational subjects).

Upper bound	Lower bound	AS estimates
0.9899	0.0000	0.6951
0.9899	0.0000	N.A.
0.9985	0.9985	0.9985
0.9993	0.9993	0.9994
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9995	0.9997
1.0000	0.9997	1.0000
1.0000	1.0000	0.9981
1.0000	1.0000	0.9989
1.0000	1.0000	0.9997
1.0000	1.0000	1.0000
1.0000	1.0000	1.1118

Table D.3: Recovered bounds for present-biasedness (29 QHD rational subjects).

Upper bound	Lower bound	AS estimates
1.0000	0.9676	0.9788
1.0000	0.9972	0.9986
1.3194	0.9500	1.0139
∞	0.9500	0.9856
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0011
∞	0.9500	1.0030
∞	0.9500	1.0030
∞	0.9500	1.0781
∞	0	N.A.
∞	0	N.A.

E. Distance Measure: A Robustness Check

In Section 5.2, we introduce a measure to characterize the distance from a given dataset to rationality, be it EDU, QHD, and so on. The ideal method for obtaining such a measure is to check all the possible sequences of dropping observations, starting from dropping 1 observation, until we can find a largest subdata that pass the test. However, exhaustive checking is computationally extremely challenging. Hence we take an alternative approach: randomly drop observations and iterate this procedure. We demonstrate that the distance measure obtained by our approach does not depend on the random procedure heavily.

We prepare 3 different sets of distance measures, each of which is obtained from 10,000 iterations, for each distance measure d'_{EDU}, d'_{QHD}, and d'_{TSU}.As we see in Figure E.1, three sets result in statistically indistinguishable distributions of distance measures: the null hypothesis of equal distribution is not rejected at all conventional levels in the two-sample Kolmogorov-Smirnov test. For the analyses in Section 5.2, we merge three sets and take the shortest path from the total of 30,000 iterations.

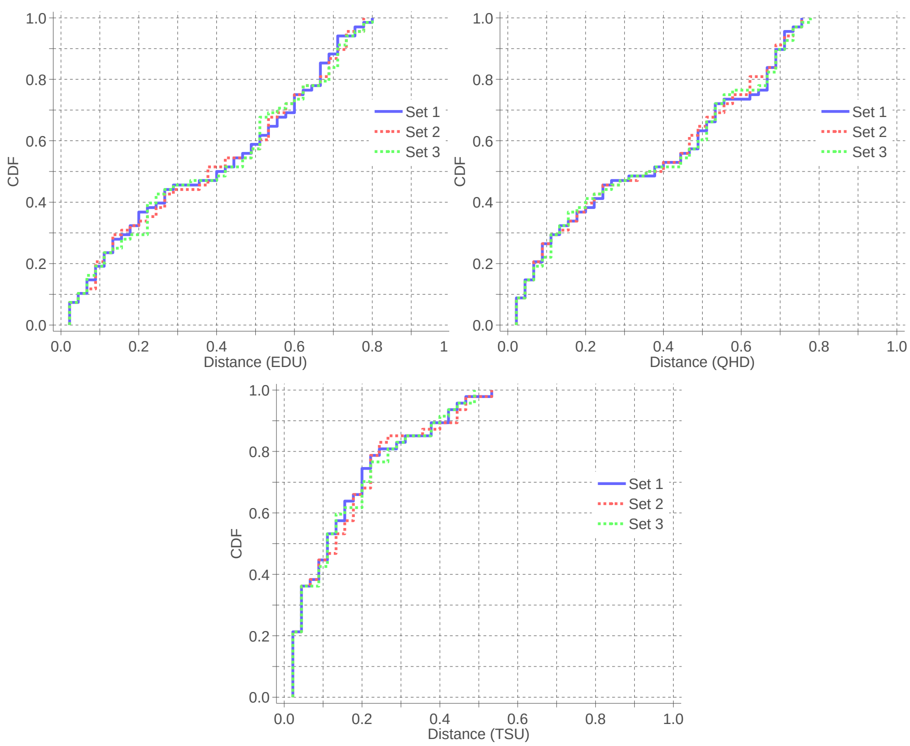


Figure E.1: Comparing the distance measures obtained from 3 sets of 10,000 iterations.

F. Jittering: Perturbing Choices

We demonstrate robustness of revealed preference tests to small perturbation in underlying preferences in Section 5.2. Here, instead of perturbing preference parameters, we add jitters on choices predicted by a QHD model with fixed set of parameters.[22]

Assume a QHD model

$$U(x_0, \dots, x_T) = \frac{1}{\alpha}x_0^\alpha + \beta \sum_{t \in T \setminus \{0\}} \delta^t \frac{1}{\alpha} x_t^\alpha$$

as in AS. For each budget in the AS experiment (there are 45 of those), the model predicts demand for sooner payment, $x(p, \tau, d; \alpha, \delta, \beta)$. We then add "jitters" to these predicted demands so that we observe $\widehat{x}(p, \tau, d; \alpha, \delta, \beta, \sigma) = x(p, \tau, d; \alpha, \delta, \beta) + \varepsilon$. Jitters are assumed to be drawn from a normal distribution, but we ensure that the jittered demand $\widehat{x}(p, \tau, d)$'s are on the budget line. In other words, jitters are drawn from a truncated normal distribution.[23]

In this exercise, we take parameters from AS aggregate estimates: $\alpha = 0.897$, $\delta = 0.999$. For the present bias parameter, we take AS aggregate estimate $\beta = 1.007$ together with other "reasonable" values such as 0.974 (aggregate estimate from Augenblick et al., forthcoming), 0.995, 1, and 1.05. As for standard deviation of the normal distribution, we use $\sigma \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1\}$.

For each set of parameters and standard deviation of white noise $(\alpha, \delta, \beta, \sigma)$, we simulate 1,000 sets of observations $\{\widehat{x}(p_b, \tau_b, d_b; \alpha, \delta, \beta, \sigma)\}_{b=1}^{45}$. We then perform our EDU and QHD tests.

Table F.1 reports pass rates for the QHD test for each set of parameters and standard deviation. When the standard deviation is $\sigma = 0.001$, the simulated dataset always pass the QHD test. As the standard deviation increases, pass rates decrease at the speed depending on the parameter configuration.[24]

Table F.2 reports the same statistics for the EDU test. A notable feature in this simulation is that the dataset generated by non-EDU preferences (i.e., $\beta = 0.995$ and 1.007) pass the EDU test in many occasions. As in the case of the QHD test, pass rates decrease at the speed depending on the parameter configuration.

[22] Andreoni et al. (2013a) introduce and discuss this way of assessing the goodness-of-fit in the context of revealed preference tests, which they call the jittering measure.

[23] Andreoni et al. (2013a) note that "truncating is known to bias the frequency of corner solutions downward". An alternative approach is "censoring," which would have a bias in the opposite direction.

[24] We also confirm that predicted choices indeed pass the QHD test in the absence of jittering (4-th column in the table).

Table F.1: QHD test pass rates.

Parameters				Standard deviation (σ)							
#	α	δ	β	0	0.001	0.005	0.010	0.050	0.100	0.500	1.000
1	0.897	0.999	0.974	1.00	1.00	0.99	0.83	0.21	0.02	0.00	0.00
2	0.897	0.999	0.995	1.00	1.00	1.00	1.00	0.47	0.16	0.00	0.00
3	0.897	0.999	1.000	1.00	1.00	1.00	1.00	0.46	0.18	0.00	0.00
4	0.897	0.999	1.007	1.00	1.00	1.00	0.98	0.30	0.10	0.00	0.00
5	0.897	0.999	1.050	1.00	1.00	1.00	0.92	0.23	0.05	0.00	0.00

Table F.2: EDU test pass rates.

Parameters				Standard deviation (σ)							
#	α	δ	β	0	0.001	0.005	0.010	0.050	0.100	0.500	1.000
1	0.897	0.999	0.974	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.897	0.999	0.995	1.00	1.00	1.00	1.00	0.47	0.16	0.00	0.00
3	0.897	0.999	1.000	1.00	1.00	1.00	1.00	0.46	0.18	0.00	0.00
4	0.897	0.999	1.007	1.00	1.00	1.00	0.96	0.25	0.09	0.00	0.00
5	0.897	0.999	1.050	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This exercise has demonstrated that our revealed preference tests detect irregularities induced by white noise, but we cannot provide a definitive answer to whether the degree of irregularities necessary to violate EDU/QHD rationality is big or small (in other words, how sensitive our tests are) because we do not have clear benchmark to compare with.

As provide standard deviation of NLS error in the aggregate estimate (corresponding to parameter set #4), which is 6.13.

Alternatively, one can use variations observed in the actual experimental data to compare with standard deviations used in this exercise. Let $x_i(p_b, \tau_b, d_b)$ denote subject i's demand for sooner payment in budget b. Then, we calculate the root mean squared error (RMSE)

$$v_i = \sqrt{\frac{1}{45}\sum_{b=1}^{45}\left(x_i(p_b, \tau_b, d_b) - x(p_b, \tau_b, d_b; \alpha, \delta, \beta)\right)^2}$$

for each subject i. Table F.3 reports summary statistics for the distribution of v_i's. It is clear that the variation of the observed data measured by RMSE is much higher than the standard deviation of white noise at which we achieve 50% pass rate for the QHD test. This may suggest that about 50% of the subjects are not rationalized by QHD model because

TABLE F.3: Distributions of v_i's.

	Parameters			Percentile						
#	α	δ	β	5-th	10-th	25-th	50-th	75-th	90-th	95-th
1	0.897	0.999	0.974	3.00	3.76	4.68	5.93	6.33	7.83	10.50
2	0.897	0.999	0.995	2.91	3.66	4.60	5.93	6.17	7.94	10.61
3	0.897	0.999	1.000	2.93	3.68	4.63	5.94	6.15	7.97	10.64
4	0.897	0.999	1.007	2.95	3.71	4.62	5.91	6.18	8.02	10.67
5	0.897	0.999	1.050	3.10	3.58	4.48	5.61	6.13	8.28	10.92

of structural irregularities rather than trembling on their choices. However, we emphasize again that we do not have clear guidance for the benchmark: we demonstrate the case of v_i's but this may not be the right one to compare with.