

# Testing axioms of revealed preference in Stata

Marcos Demetry  
Linnaeus University  
Växjö, Sweden  
marcos.demetry@lnu.se

Per Hjertstrand  
Research Institute of Industrial Economics  
Stockholm, Sweden  
per.hjertstrand@ifn.se

Matthew Polisson  
University of Bristol  
Bristol, U.K.  
matthew.polisson@bristol.ac.uk

**Abstract.** The revealed preference approach in economics is central to the empirical analysis of consumer behavior. In this article, we introduce the commands `checkax`, `aei`, and `powerps` as a bundle within the package `rpaxioms`. The first command allows a user to test whether consumer expenditure data satisfy several revealed preference axioms; the second command calculates measures of goodness of fit when the data violate these axioms; and the third command calculates power against uniformly random behavior as well as predictive success for each axiom. We illustrate the commands using individual-level experimental data and household-level aggregate consumption data.

**Keywords:** `st0673`, `rpaxioms`, `checkax`, `aei`, `powerps`, revealed preference, generalized axiom of revealed preference, Afriat efficiency index, power, predictive success

## 1 Introduction

The econometrics of consumer demand is central to economic analysis—it involves testing economic theories, making out-of-sample predictions, and drawing welfare comparisons across different environments and policy regimes. Thus, the research program on empirical consumer demand has held a central position within the economics literature for many decades (see, for example, McFadden [1974] and Deaton and Muellbauer [1980]). Much of the emphasis within this literature has been on consistently estimating consumer preferences, the inherently unobservable primitive from which tests, predictions, and welfare statements can be derived.

The analogue to this research program within the context of finite data is known as revealed preference. The revealed preference approach involves checking whether a finite set of price and demand observations made on an individual consumer is compatible with economic rationality, that is, rationalizable by some form of utility maximization. Revealed preference is fully nonparametric, in the sense that it does not impose any auxiliary functional form or distributional assumptions, only the basic primitives of utility maximization. Famously, Afriat’s (1967) theorem states that a dataset is rationalizable by the maximization of a well-behaved utility function if and only if it obeys an intuitive no-cycling condition on the data. This property is commonly referred to as the

generalized axiom of revealed preference (GARP), and there are efficient algorithms for checking this axiom empirically (Varian 1982). Other notions (including special cases) of rationalizability can also be characterized in terms of observables, and the approach has given rise to a suite of revealed preference tests in the tradition of Afriat (1967) that can be used in applied empirical work.

In general, revealed preference tests are “sharp”, in the sense that they deliver a binary response as to whether observed expenditure data are compatible with an underlying behavioral model. However, given sufficiently rich data, an outright failure of even fairly permissive notions of rationalizability should not come as unexpected, and it may well be that the data are in fact very close to rationalizability. As Varian (1990) notes, “for most purposes, ‘nearly optimizing behavior’ is just as good as ‘optimizing’ behavior.” Afriat (1973) proposes to test for nearly optimizing behavior by allowing a part of the consumer’s expenditure to be “wasted”. The fraction of expenditure that is not being wasted by the consumer is usually referred to as the efficiency level of the test. Varian’s (1982) original formulation of GARP implicitly assumes an efficiency level of 1; that is, the consumer is not allowed to waste any part of his or her expenditure.

Varian (1982) introduces a simple combinatorial algorithm to test whether consumer expenditure data obey GARP. This algorithm can be easily adapted to test GARP at any efficiency level. Our first command, **checkax**, implements Varian’s algorithm to test whether a dataset satisfies GARP at any efficiency level specified by the user. The command also allows a user to test whether the data obey the following revealed preference axioms at any efficiency level: the strong axiom of revealed preference (SARP), the weak generalized axiom of revealed preference (WGARP), the weak axiom of revealed preference (WARP), the symmetric generalized axiom of revealed preference (SGARP), the homothetic axiom of revealed preference (HARP), and cyclical monotonicity (CM). These axioms and their behavioral implications are described in detail in section 2.5.

Afriat (1973) proposes an upper limit on the efficiency level at which a dataset satisfies GARP, or the critical cost efficiency, as a measure of approximate rationalizability. Hence, this index, called the Afriat efficiency index (AEI), also known as the critical cost efficiency index (CCEI), measures the severity of violations as the minimal expenditure adjustment that is required for the data to comply with GARP. Thus, Varian (1990) interprets (and extends) this measure as a “goodness-of-fit” criterion. The approach can also be applied to other axioms, and our second command, **aei**, implements the AEI for each of the following seven axioms: GARP, SARP, WGARP, WARP, SGARP, HARP, and CM. The AEI is discussed in more detail in section 2.3.

In addition to goodness-of-fit, the outcome of a revealed preference test in many empirical applications is often reported alongside some measure of power. The power of a revealed preference test, say, for GARP, is defined as the probability of rejecting GARP, given that the data were generated by some type of “irrational” consumer behavior. Bronars (1987) proposes a power index where the irrational behavior is based on Becker’s (1962) uniformly random consumption model. Thus, for this widely used power index, the choices generated by an irrational consumer are uniformly distributed along the budgetary boundaries. Our third command, **powerps**, implements the Bronars power

index for any of the axioms above at any efficiency level. This command also reports a measure of “predictive success” originally introduced by Selten (1991) and adapted to the revealed preference framework by Beatty and Crawford (2011). This measure is motivated by the idea that if the data satisfy a given revealed preference axiom, then any robust conclusion on rationalizability should, at a minimum, require the test to have high power against uniformly random behavior. Thus, the predictive success measure combines the pass rate of the revealed preference test with Bronars power index. Power and predictive success are further discussed in section 2.4.

In summary, for applied practitioners, it is important that revealed preference methods are easily implementable and reproducible. To this end, we present the package, `rpaxioms`. We illustrate our three commands—`checkax`, `aei`, and `powerps`—on two types of data that are commonly used in empirical applications of revealed preference. First, using experimental data collected by Andreoni and Miller (2002), we test whether the social allocations selected by individual subjects are compatible with utility maximization taking several different forms. Second, using aggregate household consumption data on four food categories from Poi (2002), we test whether these data can be rationalized by preferences that are common across all households.<sup>1</sup>

## 2 Revealed preference

Suppose that there are  $T$  observations of the prices and quantities of  $K$  goods. At each observation  $t = 1, \dots, T$ , the price vector is denoted by  $\mathbf{p}^t = (p_1^t, \dots, p_K^t) \gg 0$  and the quantity bundle by  $\mathbf{x}^t = (x_1^t, \dots, x_K^t) \geq 0$ . We assume that all prices are strictly positive and that all quantities are nonnegative (note that some but not all quantities at any given observation may be equal to 0; that is, all expenditures are strictly positive). The  $T$  observations of  $(\mathbf{p}^t, \mathbf{x}^t)$  then form the finite dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$ .

### 2.1 Rationalizability

The dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  is said to be rationalizable by utility maximization if there is a utility function  $U : \mathbb{R}_+^K \rightarrow \mathbb{R}$ , such that, at every observation  $t = 1, \dots, T$ ,

$$U(\mathbf{x}^t) \geq U(\mathbf{x}) \text{ for any } \mathbf{x} \in \{\mathbf{x} \in \mathbb{R}_+^K : \mathbf{p}^t \cdot \mathbf{x} \leq \mathbf{p}^t \cdot \mathbf{x}^t\}$$

Rationalizability means that we can find a utility function defined on the consumption space that assigns (weakly) higher utility to the quantity bundle  $\mathbf{x}^t$  than to any other bundle  $\mathbf{x}$  that is affordable at the prevailing prices  $\mathbf{p}^t$ . Without any further restrictions on  $U$ , any dataset  $\mathcal{D}$  is rationalizable because  $U$  could simply assign the same level of utility to every bundle in the consumption space. For the question to be meaningful, we require further structure on the utility function  $U$ .

Afriat (1967) was the first to show that a finite dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  is rationalizable by a locally nonsatiated utility function  $U$  if and only if it obeys an intu-

---

1. Poi (2002) uses the same data to illustrate the estimation of parametric demand systems in Stata.

itive property now known as the generalized axiom of revealed preference.<sup>2</sup> A dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  obeys GARP so long as the preference cycles it reveals are weak rather than strict; that is, for any cycle represented by

$$\mathbf{p}^t \cdot \mathbf{x}^t \geq \mathbf{p}^t \cdot \mathbf{x}^u, \mathbf{p}^u \cdot \mathbf{x}^u \geq \mathbf{p}^u \cdot \mathbf{x}^v, \dots, \mathbf{p}^w \cdot \mathbf{x}^w \geq \mathbf{p}^w \cdot \mathbf{x}^t$$

the inequalities cannot be strict. The intuition of GARP as a no-cycling condition on the data ought to be strong, and it should also come as no surprise that GARP is necessary for (or implied by) the maximization of a locally nonsatiated utility function. Afriat's (1967) theorem shows that GARP is also sufficient for rationalizability by a locally nonsatiated utility function  $U$ .<sup>3</sup> The importance of the result is that GARP completely characterizes the content of utility maximization in terms of observables, and it can therefore be used as an empirical test for rationalizability.

## 2.2 Approximate rationalizability

In a sufficiently rich empirical setting, it is unlikely that any dataset is exactly rationalizable, and so we require some notion of its distance to or departure from exact rationalizability. Loosely speaking, one could think of this as allowing for “error”, which has long been the convention in the econometrics of consumer demand (see, for example, McFadden [1974] and Deaton and Muellbauer [1980]). To this end, the convention in the revealed preference literature is to accommodate error through cost inefficiency, an idea first developed by Afriat (1972, 1973).

According to Afriat's (1973) formulation, the consumer “has a definite structure of wants” and “programs at a level of cost-efficiency  $e$ ”, which is tantamount to relaxing the definition of rationalizability. Consider any efficiency level  $e \in (0, 1]$ . The dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  is said to be rationalizable at efficiency level  $e$  if there is a utility function  $U : \mathbb{R}_+^K \rightarrow \mathbb{R}$ , such that, at every observation  $t = 1, \dots, T$ ,

$$U(\mathbf{x}^t) \geq U(\mathbf{x}) \text{ for any } \mathbf{x} \in \{\mathbf{x} \in \mathbb{R}_+^K : \mathbf{p}^t \cdot \mathbf{x} \leq e \mathbf{p}^t \cdot \mathbf{x}^t\}$$

When  $e = 1$ , this definition corresponds to exact rationalizability and, for any  $e < 1$ , to approximate rationalizability. Afriat (1973) shows that a dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  is rationalizable at efficiency level  $e$  by a locally nonsatiated utility function  $U$  if and only if, for any cycle represented by

$$e \mathbf{p}^t \cdot \mathbf{x}^t \geq \mathbf{p}^t \cdot \mathbf{x}^u, e \mathbf{p}^u \cdot \mathbf{x}^u \geq \mathbf{p}^u \cdot \mathbf{x}^v, \dots, e \mathbf{p}^w \cdot \mathbf{x}^w \geq \mathbf{p}^w \cdot \mathbf{x}^t$$

the inequalities cannot be strict. The equivalent condition known as  $e$ GARP is necessary and sufficient for approximate rationalizability by a locally nonsatiated utility function  $U$ .<sup>4</sup> When  $e = 1$ ,  $e$ GARP and GARP coincide.

2. To be precise, Afriat (1967) refers to this property as cyclical consistency, which Varian (1982) shows is equivalent to GARP.

3. Furthermore, Afriat (1967) shows that such a  $U$  can be chosen to be continuous, strictly increasing, and concave; these properties, within the current context, have no empirical content. See also Diewert (1973) and Varian (1982) for proofs of this seminal result.

4. Once again,  $U$  can be chosen to be continuous, strictly increasing, and concave. See also Halevy, Persitz, and Zrill (2018) for a proof of this result.

The operationalization of *e*GARP as an empirical test is straightforward. Consider any efficiency level  $e \in (0, 1]$ . For any pair of observations  $(t, s)$ , we say that  $\mathbf{x}^t$  is directly revealed preferred to  $\mathbf{x}^s$  at efficiency level  $e$ , written  $\mathbf{x}^t R_e^D \mathbf{x}^s$ , if  $e \mathbf{p}^t \cdot \mathbf{x}^t \geq \mathbf{p}^t \cdot \mathbf{x}^s$ . This means that  $\mathbf{x}^t$  is chosen even though the cost of the bundle  $\mathbf{x}^s$  (at the prevailing prices  $\mathbf{p}^t$ ) does not exceed  $e \mathbf{p}^t \cdot \mathbf{x}^t$ . Analogously, we say that  $\mathbf{x}^t$  is strictly directly revealed preferred to  $\mathbf{x}^s$  at efficiency level  $e$ , written  $\mathbf{x}^t P_e^D \mathbf{x}^s$ , if  $e \mathbf{p}^t \cdot \mathbf{x}^t > \mathbf{p}^t \cdot \mathbf{x}^s$ . We say that  $\mathbf{x}^t$  is revealed preferred to  $\mathbf{x}^s$  at efficiency level  $e$ , written  $\mathbf{x}^t R_e \mathbf{x}^s$ , if there exists a sequence of observations  $(t, u, v, \dots, w, s)$  such that  $\mathbf{x}^t R_e^D \mathbf{x}^u$ ,  $\mathbf{x}^u R_e^D \mathbf{x}^v$ ,  $\dots$ ,  $\mathbf{x}^w R_e^D \mathbf{x}^s$ . Hence,  $R_e$  is the transitive closure of  $R_e^D$ . When  $e = 1$ , these relations reduce to the usual revealed preference relations (Varian 1982).

A dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  satisfies *e*GARP if  $\mathbf{x}^t R_e \mathbf{x}^s$  implies  $e \mathbf{p}^s \cdot \mathbf{x}^s \leq \mathbf{p}^s \cdot \mathbf{x}^t$ .

The *e*GARP condition can be tested at any efficiency level  $e$  by slightly modifying the algorithm proposed by Varian (1982). First, the relations  $R_e^D$  and  $P_e^D$  are formed by constructing the  $T \times T$  matrices  $\mathbf{R}_e^D$  and  $\mathbf{P}_e^D$ , where the  $(t, s)$ th elements of  $\mathbf{R}_e^D$  and  $\mathbf{P}_e^D$  are equal to 1 if  $e \mathbf{p}^t \cdot \mathbf{x}^t \geq \mathbf{p}^t \cdot \mathbf{x}^s$  and  $e \mathbf{p}^t \cdot \mathbf{x}^t > \mathbf{p}^t \cdot \mathbf{x}^s$  and 0 otherwise, respectively. Second, the relation  $R_e$  is formed by calculating the transitive closure of the matrix  $\mathbf{R}_e^D$ , which gives a  $T \times T$  matrix  $\mathbf{R}_e$ , where the  $(t, s)$ th element of  $\mathbf{R}_e$  is equal to 1 if  $\mathbf{x}^t R_e \mathbf{x}^s$  and 0 otherwise. Varian (1982) suggests calculating  $\mathbf{R}_e$  using Warshall's (1962) algorithm. Finally, *e*GARP is violated whenever the  $(t, s)$ th element of  $\mathbf{R}_e$  and the  $(s, t)$ th element of  $\mathbf{P}_e^D$  are both equal to 1. The total number of violations is given by the number of pairs  $(t, s)$ , with  $t \neq s$ , such that this occurs. Therefore, with  $T$  observations, the total possible number of *e*GARP violations is  $T(T-1)$ , and the fraction of violations is given by the ratio of the number of violations to  $T(T-1)$ .

Our first command, `checkax`, constructs  $\mathbf{R}_e^D$ ,  $\mathbf{P}_e^D$ , and  $\mathbf{R}_e$  at any efficiency level  $e$  specified by the user. The last is constructed using a vectorized version of Warshall's algorithm. The command `checkax` then reports to the user whether the data satisfy *e*GARP, as well as the number and fraction of violations.<sup>5</sup>

## 2.3 The AEI

It is clear that any dataset is approximately rationalizable at some sufficiently small efficiency level  $e \in (0, 1]$ . Afriat (1973) defines the CCEI or the AEI as the maximal value of  $e$  (the supremum, to be precise) such that a dataset obeys *e*GARP.<sup>6</sup> Varian (1990) interprets the AEI as a measure of “goodness of fit” in terms of wasted expenditure: if a consumer has an AEI of  $e^* < 1$ , then he or she could have obtained the same level of utility by spending only the fraction  $e^*$  of what he or she actually spent. This is the sense in which the consumer is exhibiting cost inefficiency, and in many applications,

5. Swofford and Whitney (1987) originally suggest using the number of violations as a goodness-of-fit measure, while Famulari (1995) proposes a related metric, which can roughly be interpreted as the fraction of violations.

6. Note that Varian (1990) extends Afriat's (1973) approach by allowing the amount of cost inefficiency to vary across observations, so that  $\mathbf{e} = (e^1, \dots, e^T) \in (0, 1]^T$  represents a vector of efficiency levels; Halevy, Persitz, and Zrill (2018) discuss the different aggregators of  $\mathbf{e}$ , which include the Afriat (1973), Varian (1990), and Houtman and Maks (1985) indices.

the AEI is interpreted as a measure of “rationality” or “decision-making quality” (see, for example, Choi et al. [2014]).<sup>7</sup>

Our second command, `aei`, calculates the AEI by implementing the binary search algorithm described in Varian (1990).

## 2.4 Power and predictive success

Within applications of revealed preference, alongside goodness of fit, it is important to know something about the power of the test or empirical environment. To this end, the convention in the applied revealed preference literature is to test against an alternative behavioral model, which is typically random choice and interpreted as “naive” or “irrational”. The notion of “irrationality” that underpins the Bronars (1987) power index is based on a model of uniformly random consumption, in which all feasible consumption allocations (that is, frontier bundles) are equally likely to be chosen.

Bronars (1987) suggests implementing the index using Monte Carlo methods. The first step consists of generating artificial budget shares that are consistent with uniformly random consumption. At each observation, this involves generating  $K$  random variables drawn from the Dirichlet distribution with all parameters (characterizing this distribution) set equal to 1. By construction, at each observation, these random variables are uniformly distributed on the  $(K - 1)$ -dimensional unit simplex and, consequently, can be interpreted as budget shares in the uniformly random model. The second step solves for each uniformly random consumption quantity (denoted by  $q_k^t$ ) from the budget share equation given by  $w_k^t = p_k^t q_k^t / \mathbf{p}^t \cdot \mathbf{x}^t$ , where each  $w_k^t$  denotes an artificial budget share generated in the first step. (Notice that  $\mathbf{p}^t$  and  $\mathbf{x}^t$  are given in the original dataset.) Thus, the first two steps generate a synthetic dataset across  $K$  goods and  $T$  observations that is compatible with uniformly random behavior. The third step repeats the first two steps many times and, for each repetition, checks whether the synthetic dataset of prices and uniformly random quantities satisfies *e*GARP at a given efficiency level  $e$ . The power measure is the fraction of these synthetic datasets that violates *e*GARP.

Our third command, `powerps`, calculates power at any efficiency level  $e$  and for any number of repetitions specified by the user. Moreover, the command allows the user to set the random seed in the generation of the Dirichlet random variables (in the first step of the procedure) to ensure that power calculations are perfectly replicable. The command `powerps` also reports Beatty and Crawford’s (2011) revealed preference measure of “predictive success”. For a given dataset, this measure is defined as the difference between the pass or fail indicator and one minus the Bronars power index, where the pass or fail indicator takes the value 1 if the original data obey *e*GARP at a given efficiency level  $e$  and 0 otherwise, and where the power index corresponding to *e*GARP is calculated at the same efficiency level  $e$ . This measure of predictive success can then be straightforwardly aggregated across individual datasets.

---

7. For a precise description of the relationship between (approximate) rationalizability and (approximate) cost rationalizability, which in turn motivates the AEI, see Polisson and Quah (2022).

## 2.5 Other axioms

Our commands are also implementable for other revealed preference axioms that characterize some of the common forms of utility or preference maximization. The default axiom in every command is  $e$ GARP (with  $e = 1$ ), but each command can also be executed for six other revealed preference axioms at any efficiency level  $e$  specified by the user.

- SARP. A dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  satisfies the strong axiom of revealed preference at efficiency level  $e$ , abbreviated  $e$ SARP, if  $\mathbf{x}^t R_e \mathbf{x}^s$  implies  $e \mathbf{p}^s \cdot \mathbf{x}^s < \mathbf{p}^s \cdot \mathbf{x}^t$  whenever  $\mathbf{x}^t \neq \mathbf{x}^s$ . Matzkin and Richter (1991) show that SARP ( $e = 1$ ) is necessary and sufficient for rationalizability by a continuous, strictly increasing, and strictly concave utility function. Notice that the difference between GARP and SARP is that GARP allows for “flat spots” of indifference (demand correspondences versus demand functions). Like  $e$ GARP, there can be up to  $T(T - 1)$  violations of  $e$ SARP.
- WGARP. A dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  satisfies the weak generalized axiom of revealed preference at efficiency level  $e$ , abbreviated  $e$ WGARP, if  $\mathbf{x}^t R_e^D \mathbf{x}^s$  implies  $e \mathbf{p}^s \cdot \mathbf{x}^s \leq \mathbf{p}^s \cdot \mathbf{x}^t$ . Aguiar, Hjertstrand, and Serrano (2020) show that WGARP ( $e = 1$ ) is necessary and sufficient for rationalizability by a continuous, strictly increasing, piecewise concave, and skew-symmetric preference function (see Aguiar, Hjertstrand, and Serrano [2020] for the definition of a preference function and its properties). Banerjee and Murphy (2006) show that WGARP and GARP are equivalent when  $K = 2$  (when there are two goods). The total possible number of violations of  $e$ WGARP is  $T(T - 1)/2$ .
- WARP. A dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  satisfies the weak axiom of revealed preference at efficiency level  $e$ , abbreviated  $e$ WARP, if  $\mathbf{x}^t R_e^D \mathbf{x}^s$  implies  $e \mathbf{p}^s \cdot \mathbf{x}^s < \mathbf{p}^s \cdot \mathbf{x}^t$  whenever  $\mathbf{x}^t \neq \mathbf{x}^s$ . Aguiar, Hjertstrand, and Serrano (2020) show that WARP ( $e = 1$ ) is necessary and sufficient for rationalizability by a continuous, strictly increasing, piecewise strictly concave, and skew-symmetric preference function. The difference between WARP and WGARP is analogous to the difference between SARP and GARP. Furthermore, Rose (1958) shows that WARP and SARP are equivalent when  $K = 2$ . Like  $e$ WGARP, there can be up to  $T(T - 1)/2$  violations of  $e$ WARP.
- SGARP. For any  $(t, s)$ , we can modify the definition of  $R_e^D$  so that  $\mathbf{x}^t R_e^D \mathbf{x}^s$  if  $e \mathbf{p}^t \cdot \mathbf{x}^t \geq \mathbf{p}^t \cdot \mathbf{y}^s$ , where  $\mathbf{y}^s$  is any permutation of  $\mathbf{x}^s$  and where the transitive closure  $R_e$  of  $R_e^D$  follows accordingly.<sup>8</sup> A dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  satisfies the symmetric generalized axiom of revealed preference at efficiency level  $e$ , abbreviated  $e$ SGARP, if  $\mathbf{x}^t R_e \mathbf{x}^s$  implies  $e \mathbf{p}^s \cdot \mathbf{x}^s \leq \mathbf{p}^s \cdot \mathbf{y}^t$  (where  $\mathbf{y}^t$  is any permutation of  $\mathbf{x}^t$ ). Nishimura, Ok, and Quah (2017) show that  $e$ SGARP is necessary and sufficient for rationalizability by a continuous, strictly increasing, concave, and symmetric utility function. Polisson, Quah, and Renou (2020) implement  $e$ SGARP in the

8. For example, if  $\mathbf{x}^s = (3, 1, 2)$ , then there are six permutations of  $\mathbf{x}^s$ :  $(1, 2, 3)$ ,  $(1, 3, 2)$ ,  $(2, 1, 3)$ ,  $(2, 3, 1)$ ,  $(3, 1, 2)$ , and  $(3, 2, 1)$ .

context of symmetric risk; that is, the utility function must also obey first-order stochastic dominance. The total possible number of violations of eSGARP is  $T^2$ .

- HARP. A dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  satisfies the homothetic axiom of revealed preference at efficiency level  $e$ , abbreviated eHARP, if for all distinct sequences of observations  $(s, t, u, \dots, v)$ , it must be the case that  $(\mathbf{p}^t \cdot \mathbf{x}^s)(\mathbf{p}^s \cdot \mathbf{x}^u) \dots (\mathbf{p}^v \cdot \mathbf{x}^t) \geq (e \mathbf{p}^t \cdot \mathbf{x}^t)(e \mathbf{p}^s \cdot \mathbf{x}^s) \dots (e \mathbf{p}^v \cdot \mathbf{x}^v)$ . Varian (1983) shows that HARP ( $e = 1$ ) is necessary and sufficient for rationalizability by a continuous, strictly increasing, concave, and homothetic utility function. Heufer and Hjertstrand (2019) provide a more general characterization ( $e < 1$ ) and refer to  $e^*$  in this case as the homothetic efficiency index. The command `checkax` implements eHARP as described in Varian (1983) using the Floyd–Warshall algorithm. The total possible number of violations of eHARP is  $T$ .
- CM. A dataset  $\mathcal{D} = \{(\mathbf{p}^t, \mathbf{x}^t)\}_{t=1}^T$  satisfies a cyclical monotonicity condition at efficiency level  $e$ , abbreviated eCM, if for all distinct sequences of observations  $(s, t, u, \dots, v)$ , it must be the case that  $\mathbf{p}^t \cdot (\mathbf{x}^s - e \mathbf{x}^t) + \mathbf{p}^s \cdot (\mathbf{x}^u - e \mathbf{x}^s) + \dots + \mathbf{p}^v \cdot (\mathbf{x}^t - e \mathbf{x}^v) \geq 0$ . Brown and Calsamiglia (2007) show that CM ( $e = 1$ ) is necessary and sufficient for rationalizability by a continuous, strictly increasing, concave, and quasilinear utility function. The command `checkax` implements eCM similarly to eHARP using the Floyd–Warshall algorithm. Like eHARP, there can be up to  $T$  violations of eCM.

We conclude this section with two comments.

First, notice that in general a dataset is approximately rationalizable if it could have arisen from the maximization of some utility or preference function subject to a modified budget set. Explicit theoretical support for these notions of rationalizability have been developed in the case of eGARP, eSGARP, and eHARP but not for the other axioms.

Second, we note that smoothness or differentiability has no material empirical content once cost inefficiency has been taken into account. For example, Chiappori and Rochet (1987) show that strong SARP (SSARP) is necessary and sufficient for rationalizability by an infinitely differentiable, strictly increasing, and strictly concave utility function. Suppose that a dataset obeys SARP but fails SSARP, which amounts to the same consumption bundle being chosen at two or more distinct price vectors. If we set the efficiency level to  $1 - \varepsilon$ , for some  $\varepsilon > 0$  arbitrarily small, then we could always find a smooth rationalization. Because the AEI is defined as a supremum, the AEI for SSARP would still be equal to 1. In other words, smoothness or differentiability are “untestable” in a meaningful way. See also the discussion in Polisson, Quah, and Renou (2020).



### 3 The checkax, aei, and powerps commands

All three commands take as their two main (required) arguments the  $T \times K$  price and quantity matrices:

`price(string)` specifies a  $T \times K$  price matrix, where each row corresponds to an observation  $t$  and each column to a good  $k$ . All prices are required to be strictly positive. If any of the elements in the price matrix are nonpositive (or if the price and quantity matrices have different dimensions), the commands return an error message.

`quantity(string)` specifies a  $T \times K$  quantity matrix, where each row corresponds to an observation  $t$  and each column to a good  $k$ . All quantities are required to be nonnegative. Some (but not all) quantities at a given observation may be equal to 0. If the quantity matrix violates these conditions (or if the price and quantity matrices have different dimensions), the commands return an error message.

#### 3.1 The checkax command

The syntax of `checkax` is

```
checkax, price(string) quantity(string) [axiom(string) efficiency(#)]
```

##### 3.1.1 Options for checkax

`axiom(string)` specifies the axiom or axioms that the user would like to test. The default is `axiom(eGARP)`. There are seven axioms that can be tested: `eGARP`, `eSARP`, `eWGARP`, `eWARP`, `eSGARP`, `eHARP`, and `eCM`. The user may also test all axioms simultaneously by specifying `axiom(all)`.

`efficiency(#)` specifies the efficiency level at which the user would like to test each axiom. The default is `efficiency(1)`. The efficiency level must be strictly positive and no greater than 1.

##### 3.1.2 Stored results for checkax

`checkax` stores the following in `r()`:

Scalars

<code>r(PASS_axiom)</code>	indicator for whether the data satisfy each axiom
<code>r(NUM_VIO_axiom)</code>	number of violations of each axiom
<code>r(FRAC_VIO_axiom)</code>	fraction of violations of each axiom
<code>r(GOODS)</code>	number of goods in the data
<code>r(OBS)</code>	number of observations in the data
<code>r(EFF)</code>	efficiency level at which each axiom is being tested

Macros

<code>r(AXIOM)</code>	axiom or axioms being tested
-----------------------	------------------------------

### 3.1.3 Examples of checkax

The following examples illustrate the command **checkax** using a dataset of 20 observations on the prices and quantities of 5 goods. The prices of goods 1 to 5 are **p1**, **p2**, **p3**, **p4**, and **p5**. The quantities are **x1**, **x2**, **x3**, **x4**, and **x5**. The price and quantity matrices are **P** and **X**, respectively. The first example runs **checkax** using its default options, that is, for *e*GARP at the efficiency level  $e = 1$ . The second example runs **checkax** for *e*GARP and *e*HARP at the efficiency level  $e = 0.95$ .

. use rpaxioms_example_data			
. mkmat p1-p5, matrix(P)			
. mkmat x1-x5, matrix(X)			
. checkax, price(P) quantity(X)			
	Number of obs	=	20
	Number of goods	=	5
	Efficiency level	=	1
Axiom	Pass	#vio	%vio
eGARP	0	161	.4237
. checkax, price(P) quantity(X) axiom(eGARP eHARP) efficiency(0.95)			
	Number of obs	=	20
	Number of goods	=	5
	Efficiency level	=	.95
Axiom	Pass	#vio	%vio
eGARP	0	104	.2737
eHARP	0	20	1

## 3.2 The aei command

The syntax of **aei** is

```
aei, price(string) quantity(string) [axiom(string) tolerance(#)]
```

### 3.2.1 Options for aei

**axiom**(string) specifies the axiom or axioms that the user would like to test. The default is **axiom**(eGARP). There are seven axioms that can be tested: **eGARP**, **eSARP**, **eWGARP**, **eWARP**, **eSGARP**, **eHARP**, and **eCM**. The user may also test all axioms simultaneously by specifying **axiom**(all).

**tolerance**(#) sets the tolerance level of the termination criterion  $10^{-n}$  by specifying the integer  $n$ . For example, **tolerance**(10) sets the tolerance level to  $10^{-10}$ . The default is **tolerance**(6), which gives the default tolerance level  $10^{-6}$ . The integer  $n$  in the termination criterion  $10^{-n}$  cannot be less than 1 or greater than 18.

### 3.2.2 Stored results for aei

`aei` stores the following in `r()`:

Scalars

<code>r(AEI_axiom)</code>	AEI for each axiom
<code>r(TOL)</code>	tolerance level for termination criterion
<code>r(GOODS)</code>	number of goods in the data
<code>r(OBS)</code>	number of observations in the data

Macros

<code>r(AXIOM)</code>	axiom or axioms being tested
-----------------------	------------------------------

### 3.2.3 Examples for aei

The following examples illustrate the command `aei` using the same data as above. The first example runs `aei` using its default options, that is, for `eGARP` with a tolerance level of  $10^{-6}$ . The second example runs `aei` for `eGARP` and `eHARP` with the tolerance level set to  $10^{-10}$  and shows that the command `quietly` can be used to suppress the output.

```
. aei, price(P) quantity(X)
      Number of obs      =          20
      Number of goods    =           5
      Tolerance level    =       1.0e-06
```

Axiom	AEI
eGARP	.9055848

```
. quietly aei, price(P) quantity(X) axiom(eGARP eHARP) tolerance(10)
```

## 3.3 The powerps command

The syntax of `powerps` is

```
powerps, price(string) quantity(string) [axiom(string) efficiency(#)
      simulations(#) seed(#) aei tolerance(#) progressbar]
```

### 3.3.1 Options for powerps

`axiom(string)` specifies the axiom or axioms that the user would like to test. The default is `axiom(eGARP)`. There are seven axioms that can be tested: `eGARP`, `eSARP`, `eWGARP`, `eWARP`, `eSGARP`, `eHARP`, and `eCM`. The user may also test all axioms simultaneously by specifying `axiom(all)`.

`efficiency(#)` specifies the efficiency level at which the user would like to test each axiom. The default is `efficiency(1)`. The efficiency level must be strictly positive and no greater than 1.

`simulations(#)` specifies the number of repetitions of the simulated uniformly random data. The default is `simulations(1000)`.

`seed(#)` specifies the random seed in the generation of the Dirichlet random numbers. The default is `seed(12345)`.

`aei` specifies whether the user wants to compute the AEI for each simulated dataset and specified axiom. Note that including this option may increase computation times substantially.

`tolerance(#)` sets the tolerance level of the termination criterion  $10^{-n}$  by specifying the integer  $n$  when computing the AEI. See section 3.2 for a more detailed description. This option is useful only in combination with the `aei` option.

`progressbar` specifies if the user wants to display the number of repetitions that have been executed.

### 3.3.2 Stored results for `powerps`

`powerps` stores the following in `r()`:

#### Scalars

<code>r(POWER_</code> <i>axiom</i> <code>)</code>	computed power for each axiom
<code>r(PS_</code> <i>axiom</i> <code>)</code>	computed predictive success for each axiom
<code>r(PASS_</code> <i>axiom</i> <code>)</code>	indicator for whether the data satisfy each axiom
<code>r(AEI_</code> <i>axiom</i> <code>)</code>	AEI for each axiom
<code>r(SIM)</code>	number of repetitions in the simulated uniformly random data
<code>r(TOL)</code>	tolerance level for termination criterion (if <code>aei</code> option specified)
<code>r(EFF)</code>	efficiency level at which each axiom is being tested
<code>r(GOODS)</code>	number of goods in the data
<code>r(OBS)</code>	number of observations in the data

#### Macros

<code>r(AXIOM)</code>	axiom or axioms being tested
-----------------------	------------------------------

#### Matrices

<code>r(SUMSTATS_</code> <i>axiom</i> <code>)</code>	summary statistics for random data: number of violations, fraction of violations, and AEI (if <code>aei</code> option specified)
<code>r(SIMRESULTS_</code> <i>axiom</i> <code>)</code>	number of violations, fraction of violations, and AEI (if <code>aei</code> option specified) for every simulated uniformly random dataset

### 3.3.3 Examples for `powerps`

The following examples illustrate the command `powerps` using the same data as above. The first example runs `powerps` for the axioms eGARP and eHARP. All other options are set to their defaults. The second example also runs `powerps` for the axioms eGARP and eHARP but now includes the option `aei`, which calculates the AEI for both axioms for each of the 1,000 simulated datasets. Note that including the `aei` option increases computation time substantially.

```
. powerps, price(P) quantity(X) axiom(eGARP eHARP)
      Number of obs      =      20
      Number of goods    =       5
      Simulations        =     1000
      Efficiency level    =       1
```

Axioms	Power	PS	Pass	AEI
eGARP	.995	-.005	0	.9055848
eHARP	1	0	0	.8449683

Summary statistics for simulations:

eGARP	#vio	%vio
Mean	47.339	.1245762
Std. Dev.	29.45589	.0775135
Min	0	0
Q1	24	.0632
Median	45	.1184
Q3	68.5	.18025
Max	143	.3763

eHARP	#vio	%vio
Mean	20	1
Std. Dev.	0	0
Min	20	1
Q1	20	1
Median	20	1
Q3	20	1
Max	20	1

```
. powerps, price(P) quantity(X) axiom(eGARP eHARP) aei
      Number of obs      =      20
      Number of goods    =       5
      Simulations        =     1000
      Efficiency level    =       1
```

Axioms	Power	PS	Pass	AEI
eGARP	.995	-.005	0	.9055848
eHARP	1	0	0	.8449683

Summary statistics for simulations:

eGARP	#vio	%vio	AEI
Mean	47.339	.1245762	.842074
Std. Dev.	29.45589	.0775135	.0814885
Min	0	0	.5616641
Q1	24	.0632	.7924724
Median	45	.1184	.8516641
Q3	68.5	.18025	.9015746
Max	143	.3763	1

eHARP	#vio	%vio	AEI
Mean	20	1	.7268926
Std. Dev.	0	0	.0760639
Min	20	1	.4819741
Q1	20	1	.6767941
Median	20	1	.7307339
Q3	20	1	.7845821
Max	20	1	.8955998

## 4 Empirical illustrations

This section illustrates how to implement our commands using two types of data that are common in many revealed preference applications. The first dataset contains the individual choices of experimental subjects. Such controlled environments are desirable from the perspective of empirical testing because relative prices can be randomized or calibrated across observations to engineer a sufficiently powerful test of, say, utility maximization. In our empirical illustration, we analyze the budgetary data collected in Andreoni and Miller (2002); other prominent examples of experiments involving budgetary designs include Choi et al. (2007, 2014); Andreoni and Sprenger (2012); and Halevy, Persitz, and Zrill (2018). The second dataset contains annual household food consumption within broad categories. Aggregate household-level data have long been used to estimate parametric demand systems (see, for example, Deaton and Muellbauer [1980]; Banks, Blundell, and Lewbel [1997]; and Lewbel and Pendakur [2009]); and moreover, Poi (2002) makes use of the same dataset to illustrate the estimation of parametric demand systems in Stata.

### 4.1 Experimental data

Andreoni and Miller (2002) test whether the social choices of experimental subjects are rational, using a dictator game in which one subject (the dictator) allocates token endowments between himself and another subject (the beneficiary) according to some rate of transfer. The payoffs of the dictator and the beneficiary are essentially two distinct goods, and the transfer rates are the price ratios. The experiment contains 2 parts, where 142 subjects (group 1) face  $T = 8$  decision rounds, and where 34 subjects (group 2) face  $T = 11$  rounds. In this illustration, we focus on subjects in group 1.

Andreoni and Miller (2002) find that 13 subjects in group 1 violate rationality and, for each of these 13 subjects, report the AEI (for  $e$ GARP) and the number of violations of  $e$ GARP,  $e$ SARP, and  $e$ WARP at the efficiency level  $e = 1$  (see table II in Andreoni and Miller [2002]). Banerjee and Murphy (2006) complement this analysis and report the number of violations of  $e$ WGARP at the efficiency level  $e = 1$  (see table 1 in Banerjee and Murphy [2006]). Using the commands `checkax` and `aei`, the following code replicates these results:

```
. local axioms eGARP eWGARP eSARP eWARP
. forvalues subject = 1/142 {
2. foreach axiom of local axioms {
3.   checkax, price(P) quantity(Q`subject') axiom(`axiom')
4. }
5. aei, price(P) quantity(Q`subject') axiom(eGARP)
6. }
(output omitted)
```

The results from the preceding code are reported in table 1.

Table 1. Replication of results in Andreoni and Miller (2002, table II) and Banerjee and Murphy (2006, table 1)\*

Subject	Number (fraction) of violations				AEI ( $e$ GARP)
	$e$ GARP	$e$ WGARP	$e$ SARP	$e$ WARP	
3	2 (0.036)	1 (0.036)	4 (0.071)	1 (0.036)	1.000 <sup>†</sup>
38	8 (0.143)	2 (0.071)	9 (0.161)	2 (0.071)	0.917
40	8 (0.143)	3 (0.107)	11 (0.196)	3 (0.107)	0.833
41	1 (0.018)	1 (0.036)	2 (0.036)	1 (0.036)	1.000 <sup>†</sup>
47	1 (0.018)	1 (0.036)	2 (0.036)	1 (0.036)	1.000 <sup>†</sup>
61	4 (0.071)	1 (0.036)	5 (0.089)	1 (0.036)	0.917
72	1 (0.018)	1 (0.036)	2 (0.036)	1 (0.036)	1.000 <sup>†</sup>
87	1 (0.018)	1 (0.036)	2 (0.036)	1 (0.036)	1.000 <sup>†</sup>
90	2 (0.036)	1 (0.036)	2 (0.036)	1 (0.036)	0.975
104	1 (0.018)	1 (0.036)	3 (0.054)	1 (0.036)	1.000 <sup>†</sup>
126	1 (0.018)	1 (0.036)	4 (0.071)	1 (0.036)	1.000 <sup>†</sup>
137	1 (0.018)	1 (0.036)	2 (0.036)	1 (0.036)	1.000 <sup>†</sup>
139	1 (0.018)	1 (0.036)	2 (0.036)	1 (0.036)	1.000 <sup>†</sup>

\*The number (and fraction) of violations is reported at the efficiency level  $e = 1$ .

<sup>†</sup>The symbol (<sup>†</sup>) indicates that an  $\varepsilon$ -change in choices eliminates all GARP violations.

In figure 1, we plot the fraction of the 142 subjects satisfying  $e$ GARP,  $e$ SGARP,  $e$ HARP, and  $e$ CM for values of  $e$  between 0.85 and 1 in an equally spaced grid with increments of 0.01. The results used to generate figure 1 are obtained by looping over all subjects, axioms, and efficiency levels in the grid and evaluating the command `checkax` for each subject, axiom, and efficiency level. The following line of code illustrates one such evaluation:

```
. checkax, price(P) quantity(Q`subject') efficiency(0.85)
(output omitted)
```

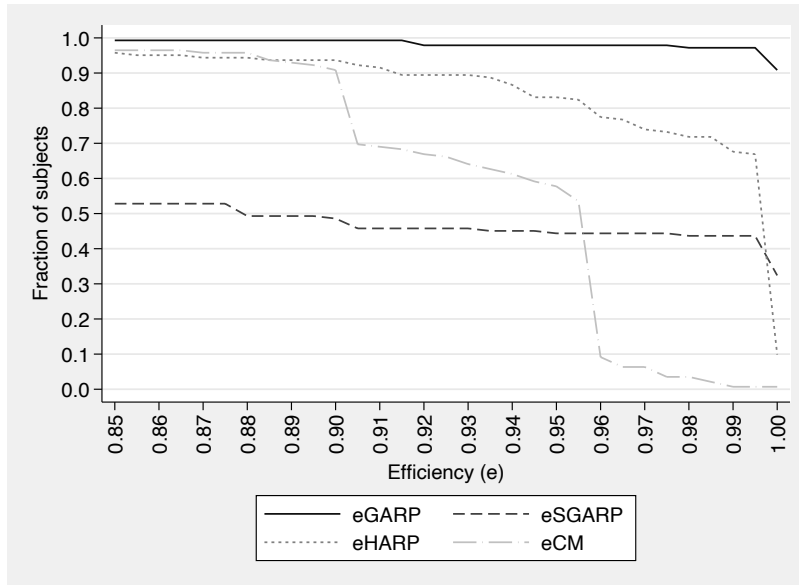


Figure 1. AEI distributions for  $e$ GARP,  $e$ SGARP,  $e$ HARP, and  $e$ CM

Because subjects are choosing from among bundles of two goods,  $e$ GARP ( $e$ SARP) and  $e$ WGARP ( $e$ WARP) are equivalent and must by construction deliver identical results in terms of pass rates (but not in terms of the number and fraction of violations). Furthermore, while theoretically possible, the empirical differences between  $e$ GARP ( $e$ WGARP) and  $e$ SARP ( $e$ WARP) are negligible, implying that distinctions between demand correspondences and demand functions are not of first-order importance within these data. Because neither Andreoni and Miller (2002) nor Banerjee and Murphy (2006) report any results for  $e$ SGARP,  $e$ HARP, or  $e$ CM, we give these axioms more attention: we calculate the mean, standard deviation, minimum, first quartile (Q1), median, third quartile (Q3), and maximum of the number (and fraction) of violations and of the AEIs corresponding to  $e$ SGARP,  $e$ HARP, and  $e$ CM. The results are displayed in table 2.



Table 2. Summary statistics for *e*SGARP, *e*HARP, and *e*CM\*

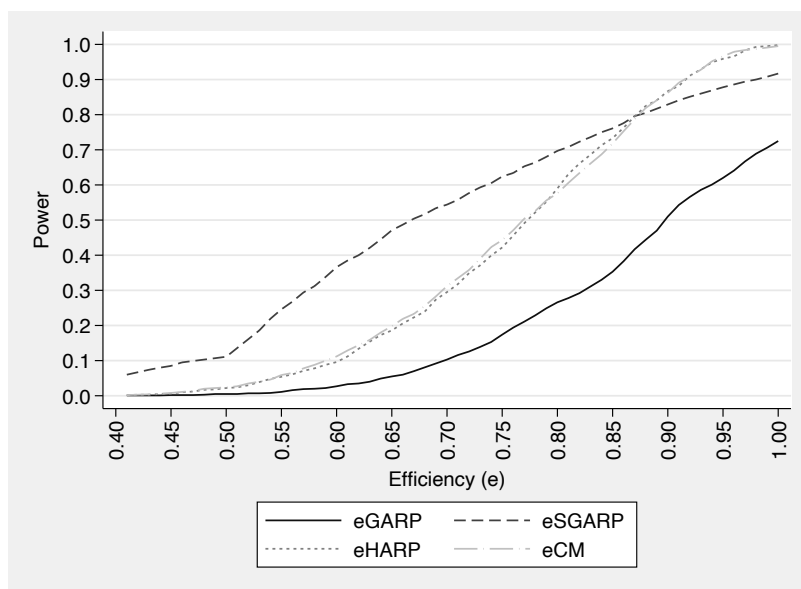
Statistic	Number (fraction) of violations			AEI		
	<i>e</i> SGARP	<i>e</i> HARP	<i>e</i> CM	<i>e</i> SGARP	<i>e</i> HARP	<i>e</i> CM
Mean	16.47 (0.257)	6.29 (0.789)	7.68 (0.960)	0.745	0.976	0.935
Std. dev.	16.80 (0.262)	2.90 (0.362)	1.03 (0.129)	0.288	0.049	0.035
Minimum	0 (0.000)	0 (0.000)	0 (0.000)	0.333	0.707	0.800
Q1	0 (0.000)	5 (0.625)	8 (1.000)	0.333	0.966	0.905
Median	8 (0.125)	8 (1.000)	8 (1.000)	0.875	1.000	0.957
Q3	37 (0.578)	8 (1.000)	8 (1.000)	1.000	1.000	0.957
Maximum	41 (0.641)	8 (1.000)	8 (1.000)	1.000	1.000	1.000

\*The number (and fraction) of violations is reported at the efficiency level  $e = 1$ .

Finally, we turn to power and predictive success. By looping over all subjects, axioms, and values of  $e$  between 0.4 and 1.0, we calculate the power and predictive success for every subject, axiom, and efficiency level in the grid. The following line of code illustrates one such evaluation:

```
. powerps, price(P) quantity(Q`subject') efficiency(0.4)
(output omitted)
```

We summarize the results in three different ways. First, figure 2 plots the power of *e*GARP, *e*SGARP, *e*HARP, and *e*CM for every efficiency level in the grid. Note that because all subjects face the same budgets, the power of each test is identical across subjects. Second, table 3 gives the mean, standard deviation, minimum, first quartile (Q1), median, third quartile (Q3), and maximum of the number (and fraction) of violations and of the AEIs corresponding to *e*SGARP, *e*HARP, and *e*CM, over all repetitions in the simulated uniformly random data. Third, figure 3(a) plots the mean predictive success across all subjects at each efficiency level in the grid, and figure 3(b) is a subject-level scatterplot of *e*HARP versus *e*GARP at selected efficiency levels.

Figure 2. Power of  $e$ GARP,  $e$ SGARP,  $e$ HARP, and  $e$ CMTable 3. Power summary statistics for  $e$ SGARP,  $e$ HARP, and  $e$ CM\*

Statistic	Number (fraction) of violations			AEI		
	$e$ SGARP	$e$ HARP	$e$ CM	$e$ SGARP	$e$ HARP	$e$ CM
Mean	17.53 (0.274)	7.96 (0.995)	7.93 (0.992)	0.693	0.763	0.761
Std. dev.	12.29 (0.192)	0.47 (0.058)	0.65 (0.081)	0.181	0.120	0.124
Minimum	0 (0.000)	0 (0.000)	0 (0.000)	0.335	0.358	0.358
Q1	8 (0.125)	8 (1.000)	8 (1.000)	0.551	0.684	0.675
Median	15 (0.234)	8 (1.000)	8 (1.000)	0.667	0.773	0.769
Q3	27 (0.422)	8 (1.000)	8 (1.000)	0.840	0.856	0.859
Maximum	53 (0.828)	8 (1.000)	8 (1.000)	1.000	1.000	1.000

\*The number (and fraction) of violations is reported at the efficiency level  $e = 1$ .

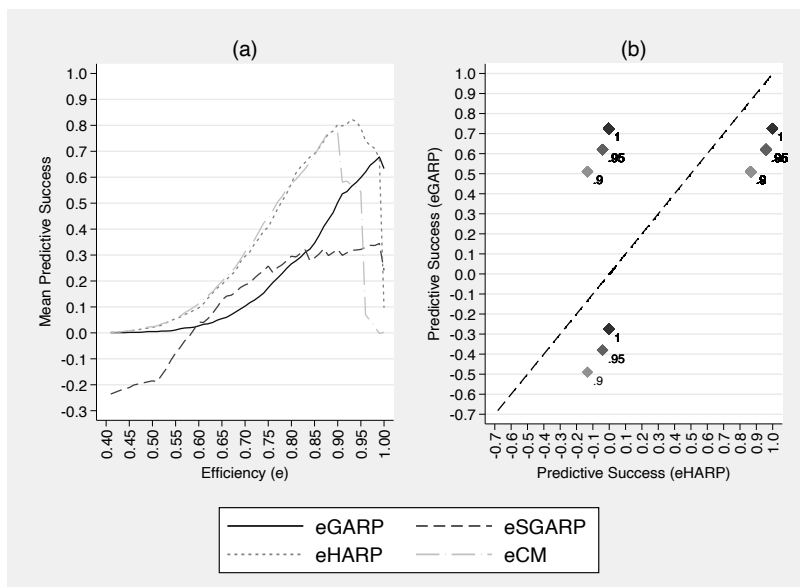


Figure 3. (a) Mean predictive success for  $eGARP$ ,  $eSGARP$ ,  $eHARP$ , and  $eCM$ ; (b) scatterplot of  $eHARP$  versus  $eGARP$ . In panel (b), the dashed line is the  $45^\circ$  line, and the marker numbers refer to efficiency levels.

## 4.2 Aggregate household consumption data

In the second empirical illustration, we use aggregate household consumption data from the 1987–1988 Nationwide Food Consumption Survey conducted by the United States Department of Agriculture. This dataset is used by Poi (2002) to illustrate how Stata's `ml` command can be used to fit the quadratic almost-ideal demand system. This dataset is named `food.dta` in the repository, *Datasets for Stata Base Reference Manual, Release 17* (<https://www.stata-press.com/data/r17/r.html>), and contains budget shares and prices for the following four aggregated food categories: meats, fruits and vegetables, breads and cereals, and miscellaneous. As in Poi (2002), we use a sample of 4,048 households.

To test whether the data can be rationalized by preferences that are common across all households, we compute the AEI for GARP and WGARP.

```

. use http://www.stata-press.com/data/r17/food, clear
(Four food groups)
. mkmatrix p1 p2 p3 p4, matrix(P)
. forvalues i = 1(1)4 {
  2. generate x`i' = w`i'* expfd/p`i'
  3. }
. mkmatrix x1 x2 x3 x4, matrix(X)
. aei, price(P) quantity(X)

      Number of obs      =      4048
      Number of goods    =         4
      Tolerance level    =     1.0e-06

```

Axiom	AEI
eGARP	.459821

```

. aei, price(P) quantity(X) axiom(eWGARP)

      Number of obs      =      4048
      Number of goods    =         4
      Tolerance level    =     1.0e-06

```

Axiom	AEI
eWGARP	.459821

We find that testing for *e*GARP takes considerably longer than testing for *e*WGARP, which suggests that the main computational burden in testing for *e*GARP is associated with the calculation of the transitive closure of the revealed preference relation. Interestingly, we find identical values of the AEI for GARP and WGARP, suggesting that none of the violations of GARP can be attributed to violations of transitivity.

Finally, because *e*WGARP is considerably faster to test than *e*GARP, we calculate the power of *e*WGARP at an efficiency level equal to the AEI for WGARP.

```

. aei, price(P) quantity(X) axiom(eWGARP)
(output omitted)
. return list
scalars:
      r(TOL) = 1.000000000000e-06
      r(GOODS) = 4
      r(OBS) = 4048
      r(AEI_eWGARP) = .4598209857940674
macros:
      r(AXIOM) : " eWGARP"

```

```
. powerps, price(P) quantity(X) axiom(eWGARP) efficiency(`r(AEI_eWGARP)')
                                Number of obs      =      4048
                                Number of goods      =         4
                                Simulations           =     1000
                                Efficiency level      =      .46
```

Axioms	Power	PS	Pass	AEI
eWGARP	.423	.423	1	.459821

Summary statistics for simulations:

eWGARP	#vio	%vio
Mean	.832	0
Std. Dev.	1.790246	0
Min	0	0
Q1	0	0
Median	0	0
Q3	1	0
Max	39	0

## 5 Conclusions

In this article, we presented the new commands `checkax`, `aei`, and `powerps` to test whether observed data on prices and quantities can be rationalized by different notions of utility maximization. The commands are implementations of nonparametric revealed preference restrictions that can be formulated as combinatorial algorithms. An important property of such algorithms is that they converge in a finite number of steps and, consequently, can be implemented on rather large datasets. Although the commands are implementations of characterizations that are intrinsically deterministic (in the sense that they lack stochastic components), they also allow the user to calculate diagnostic measures such as goodness of fit, power, and predictive success of the underlying behavioral model.

The package `rpaxioms` contains implementations of perhaps the most basic concepts in the empirical revealed preference literature. Two natural extensions for future work come to mind. The first is to provide implementations of revealed preference characterizations of other behavioral models, including special cases on preferences which amount to different forms of separability (for example, expected utility under risk and exponential discounted utility over time). The second is to provide implementations of more disaggregated measures of goodness of fit and power. Although some of these models and measures can be implemented by solving (mixed-integer) linear programming problems, this is not a trivial task, and the computational complexity of doing so crucially depends on the algorithms used to solve such problems.

## 6 Acknowledgments

We thank the editor Stephen Jenkins, a reviewer, Glenn Nielsen, and John Quah for helpful comments and useful suggestions. Per Hjertstrand thanks Torsten Söderbergs stiftelse for financial support.

## 7 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 22-2
. net install st0673      (to install program files, if available)
. net get st0673          (to install ancillary files, if available)
```

To install `checkax`, `aei`, and `powerps` and obtain the data from the Statistical Software Components Archive, type

```
. ssc install rpaxioms
. net get rpaxioms
```

Note that typing `net get` downloads the dataset and do-file to your current working directory.

## 8 References

- Afriat, S. N. 1967. The construction of utility functions from expenditure data. *International Economic Review* 8: 67–77. <https://doi.org/10.2307/2525382>.
- . 1972. Efficiency estimation of production functions. *International Economic Review* 13: 568–598. <https://doi.org/10.2307/2525845>.
- . 1973. On a system of inequalities in demand analysis: An extension of the classical method. *International Economic Review* 14: 460–472. <https://doi.org/10.2307/2525934>.
- Aguiar, V. H., P. Hjertstrand, and R. Serrano. 2020. A rationalization of the weak axiom of revealed preference. Bravo Working Paper 2020-016, Brown University and Research Institute of Industrial Economics (IFN) Working Paper No. 1321. <https://dx.doi.org/10.2139/ssrn.3543674>.
- Andreoni, J., and J. Miller. 2002. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* 70: 737–753. <https://doi.org/10.1111/1468-0262.00302>.
- Andreoni, J., and C. Sprenger. 2012. Estimating time preferences from convex budgets. *American Economic Review* 102: 3333–3356. <https://doi.org/10.1257/aer.102.7.3333>.

- Banerjee, S., and J. H. Murphy. 2006. A simplified test for preference rationality of two-commodity choice. *Experimental Economics* 9: 67–75. <https://doi.org/10.1007/s10683-006-4313-6>.
- Banks, J., R. Blundell, and A. Lewbel. 1997. Quadratic Engel curves and consumer demand. *Review of Economics and Statistics* 79: 527–539. <https://doi.org/10.1162/003465397557015>.
- Beatty, T. K. M., and I. A. Crawford. 2011. How demanding is the revealed preference approach to demand? *American Economic Review* 101: 2782–2795. <https://doi.org/10.1257/aer.101.6.2782>.
- Becker, G. S. 1962. Irrational behavior and economic theory. *Journal of Political Economy* 70: 1–13. <https://doi.org/10.1086/258584>.
- Bronars, S. G. 1987. The power of nonparametric tests of preference maximization. *Econometrica* 55: 693–698. <https://doi.org/10.2307/1913608>.
- Brown, D. J., and C. Calsamiglia. 2007. The nonparametric approach to applied welfare analysis. *Economic Theory* 31: 183–188. <https://doi.org/10.1007/s00199-006-0087-5>.
- Chiappori, P.-A., and J.-C. Rochet. 1987. Revealed preferences and differentiable demand. *Econometrica* 55: 687–691. <https://doi.org/10.2307/1913607>.
- Choi, S., R. Fisman, D. Gale, and S. Kariv. 2007. Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review* 97: 1921–1938. <https://doi.org/10.1257/aer.97.5.1921>.
- Choi, S., S. Kariv, W. Muller, and D. Silverman. 2014. Who is (more) rational? *American Economic Review* 104: 1518–1550. <https://doi.org/10.1257/aer.104.6.1518>.
- Deaton, A. S., and J. Muellbauer. 1980. An almost ideal demand system. *American Economic Review* 70: 312–326.
- Diewert, W. E. 1973. Afriat and revealed preference theory. *Review of Economic Studies* 40: 419–425. <https://doi.org/10.2307/2296461>.
- Famulari, M. 1995. A household-based, nonparametric test of demand theory. *Review of Economics and Statistics* 77: 372–382. <https://doi.org/10.2307/2109872>.
- Halevy, Y., D. Persitz, and L. Zrill. 2018. Parametric recoverability of preferences. *Journal of Political Economy* 126: 1558–1593. <https://doi.org/10.1086/697741>.
- Heufer, J., and P. Hjertstrand. 2019. Homothetic efficiency: Theory and applications. *Journal of Business and Economic Statistics* 37: 235–247. <https://doi.org/10.1080/07350015.2017.1319372>.
- Houtman, M., and J. A. H. Maks. 1985. Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve Methoden* 6: 89–104.

- Lewbel, A., and K. Pendakur. 2009. Tricks with Hicks: The EASI demand system. *American Economic Review* 99: 827–863. <https://doi.org/10.1257/aer.99.3.827>.
- Matzkin, R. L., and M. K. Richter. 1991. Testing strictly concave rationality. *Journal of Economic Theory* 53: 287–303. [https://doi.org/10.1016/0022-0531\(91\)90157-Y](https://doi.org/10.1016/0022-0531(91)90157-Y).
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, ed. P. Zarembka, 105–142. New York: Academic Press.
- Nishimura, H., E. A. Ok, and J. K.-H. Quah. 2017. A comprehensive approach to revealed preference theory. *American Economic Review* 107: 1239–1263. <https://doi.org/10.1257/aer.20150947>.
- Poi, B. P. 2002. From the help desk: Demand system estimation. *Stata Journal* 2: 403–410. <https://doi.org/10.1177/1536867X0200200406>.
- Polisson, M., and J. K.-H. Quah. 2022. Rationalizability, cost-rationalizability, and Afriat’s efficiency index. Working Paper 22/04, Institute for Fiscal Studies (IFS). <https://doi.org/10.1920/wp.ifs.2022.0422>.
- Polisson, M., J. K.-H. Quah, and L. Renou. 2020. Revealed preferences over risk and uncertainty. *American Economic Review* 110: 1782–1820. <https://doi.org/10.1257/aer.20180210>.
- Rose, H. 1958. Consistency of preference: The two-commodity case. *Review of Economic Studies* 25: 124–125. <https://doi.org/10.2307/2296210>.
- Selten, R. 1991. Properties of a measure of predictive success. *Mathematical Social Sciences* 21: 153–167. [https://doi.org/10.1016/0165-4896\(91\)90076-4](https://doi.org/10.1016/0165-4896(91)90076-4).
- Swofford, J. L., and G. A. Whitney. 1987. Nonparametric tests of utility maximization and weak separability for consumption, leisure and money. *Review of Economics and Statistics* 69: 458–464. <https://doi.org/10.2307/1925533>.
- Varian, H. R. 1982. The nonparametric approach to demand analysis. *Econometrica* 50: 945–973. <https://doi.org/10.2307/1912771>.
- . 1983. Non-parametric tests of consumer behavior. *Review of Economic Studies* 50: 99–110. <https://doi.org/10.2307/2296957>.
- . 1990. Goodness-of-fit in optimizing models. *Journal of Econometrics* 46: 125–140. [https://doi.org/10.1016/0304-4076\(90\)90051-T](https://doi.org/10.1016/0304-4076(90)90051-T).
- Warshall, S. 1962. A theorem on Boolean matrices. *Journal of the American Association of Computing Machinery* 9: 11–12. <https://doi.org/10.1145/321105.321107>.



**About the authors**

Marcos Demetry is a PhD student at Linnaeus University in Växjö, Sweden, and an affiliated doctoral student at the Research Institute of Industrial Economics (IFN) in Stockholm, Sweden.

Per Hjertstrand is an associate professor and research fellow at the Research Institute of Industrial Economics (IFN) in Stockholm, Sweden.

Matthew Polisson is a senior lecturer at the University of Bristol, United Kingdom, and a research fellow at the Institute for Fiscal Studies (IFS) in London, United Kingdom.