IFN Working Paper No. 1309, 2019

Power Against Random Expenditure Allocation for Revealed Preference Tests

Per Hjertstrand

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Power Against Random Expenditure Allocation for Revealed Preference Tests

Per Hjertstrand*

Research Institute of Industrial Economics (IFN)

Stockholm

Email: per.hjertstrand@ifn.se

April 2021

Abstract

This paper proposes new power indices for revealed preference tests. The indices are based on models of irrational consumption behavior where the consumer randomly allocates a certain fraction of expenditure. The methods allow a researcher to trace out the entire power curve against random expenditure allocation. The power indices are illustrated on altruistic choices in experimental data.

JEL Classification: C14; D12
Keywords: GARP; Power; Revealed preference

1 Introduction

The standard way of testing whether consumer choice data is consistent with rationality, i.e., the hypothesis of utility maximization, is to apply revealed preference tests. These procedures test whether the data satisfies certain axioms, such as the generalized axiom of revealed preference (GARP). If the data satisfies GARP, then there exists a continuous, strictly increasing and concave utility function rationalizing the data (Varian, 1982).

However, when GARP is violated it does not tell us how well the utility maximization model fits the data, which is usually referred to as goodness-of-fit.[1] Conversely, when GARP is satisfied it does not tell us how stringent the test is, that is, if GARP has enough bite to reject rationality if the data was generated by some type of irrational behavior. This is commonly referred to as the *power of the test*.

For this reason, empirical studies most often accompany the GARP test with diagnostic measures, such as goodness-of-fit and power, to assess how well the test performs. This paper proposes simple procedures to calculate the power against random expenditure allocation. A novel feature of these

*I thank the editor, two reviewers and Barry Jones for very useful comments. Funding from Jan Wallander och Tom Hedelius stiftelse and Marianne och Marcus Wallenberg stiftelse is gratefully acknowledged. Correspondence to: Per Hjertstrand, Research Institute of Industrial Economics (IFN), P.O. Box 55665, SE-102 15 Stockholm, Sweden. E-mail: `Per.Hjertstrand@ifn.se`.

[1] Several goodness-of-fit measures has been proposed in the literature. The most prominent is the Afriat efficiency index proposed by Afriat (1972) and Varian (1990). For other measures, see Houtman and Maks (1985) and more recently, Echenique et al. (2012) and Dean and Martin (2016). Varian (1985) proposed to deal with violations of GARP by testing whether these can be attributed to measurement errors in the data (See Jones and Edgerton, 2009, and Demuynck and Hjertstrand, 2019, for surveys).

procedures is that it allows a researcher to trace out the entire power curve against random expenditure allocation.

The procedures are generalizations of the uniformly random power index proposed by Bronars (1987). This index is based on Becker's (1962) model of irrationality, where consumption choices are uniformly distributed on the frontier of the budget set.[2] Bronars' uniformly random power index is based on the assumption that the consumer randomly allocates the entire expenditure among the goods. However, this type of irrational behavior has been criticized as being too naive, and consequently, not a very appealing alternative hypothesis to rationality (Russell et al., 1998; Varian, 2006; Andreoni et al., 2013). Nevertheless, as Varian (2006, p.105) notes, there seem to be few alternative hypotheses other than uniformly random behavior that can be applied using the same sorts of data used for revealed preference analysis. For that reason, Bronars' index has been widely used in empirical applications and is unequivocally the most popular power index in the literature.[3]

Instead, I propose a notion of irrational behavior where the consumer only randomly allocates a certain *fraction* of the expenditure. This corresponds to a model of irrationality given by the weighted average of the observed data and Bronars' purely uniformly random data. The weight can be interpreted as the fraction of expenditure that is being randomly allocated. I propose a new power index, called the partial uniform random power (PURP) index, based on this model of irrational behavior. By varying the weight and, consequently, the fraction of expenditure that is randomly allocated, the model can be used to trace out the entire power curve against uniformly random expenditure allocation.

An advantage of the PURP index is that it takes into account the information contained in the observed choices, which is not the case for Bronars' uniformly random index. On this issue, Andreoni et al. (2013, p.14-15) writes (brackets added):

> "A disadvantage is that the alternative hypothesis [i.e., Bronars uniformly random hypothesis] is perhaps too unconditional and takes no advantage of the information in observed choices about the distribution over behavior. Suppose, for instance, the budgets offered did not intersect near the points where individuals are actually choosing. Then if preferences do not conform to utility maximization, the test would be unlikely to discover it. This is true even if Bronars' analysis shows that randomly made choices provide a high probability of violations. What would be preferred, though, is an index of power based on an alternative that takes account of the choices exhibited."

The PURP index addresses this issue in a flexible way, since the fraction of expenditure that is randomly allocated is set by the researcher. If this fraction is low then the index is "very conditional" in the sense that much of the information contained in the observed choices is accounted for. Conversely, if the fraction is high the index is "very unconditional" since little of the information contained in the observed choices is taken into account.

Some other advantages of the PURP index are that it doesn't require a more detailed model of

[2] Bronars (1987) calls his uniformly random power index Algorithm 1.

[3] Andreoni et al. (2013) propose three alternative power indices. Their Afriat power index is a measure of how much budget sets would need to be shifted outward in order to generate a violation of GARP. The optimal placement index is a measure of how well an experimental design (i.e., the exogeneous choice of prices and expenditure in the data) performed relative to the best possible design that could have been generated ex post. The jittering index is a measure of how noisy the data must be in order for an experimental design to have any power. Heufer (2014) propose a conditional index, which calculates the power against uniformly random behavior, where GARP is a necessary but not sufficient condition (e.g., homotheticity and separability); See also Demuynck (2020).

behavioral bias that may be context or setting specific, and that it can be applied to a wide range of axioms and choice settings (e.g., budget choice and menu choice).

A potentially unappealing property of the PURP index is that it generates budget shares that are not centered on the actual (observed) budget shares. I therefore introduce a generalization of the PURP index, called the centered partial uniform random power (cPURP) index, where budget shares are centered on the observed shares. Thus, while the cPURP index still is based on the notion that the consumer only randomly allocates a certain fraction of the expenditure, it differs from the PURP index in the sense that the expected value of the budget shares are equal to the observed shares. As such, the cPURP index relates to another power index proposed by Bronars (1987), which also generates shares that are centered on the observed shares.[4]

Revealed preference procedures are often used in experimental economics to test whether the choices of subjects are rational. Andreoni and Miller (2002) and Fisman, Kariv and Markovits (2007) analyzed if altruistic choices of subjects are consistent with revealed preference in experimental frameworks. I illustrate the PURP and cPURP indices to these data, and find that the power may be sensitive to the amount of random expenditure allocation for many subjects in these two data sets.

This paper is organized as follows. Section 2 briefly recalls GARP from Varian (1982). Sections 3 and 4 introduces the PURP and cPURP indices. Section 5 contains the empirical application and Section 6 concludes.

2 The generalized axiom of revealed preference

Suppose a consumer chooses from $K \geq 2$ goods observed in $T \geq 2$ time periods. The goods and time periods are indexed by $\mathbb{K} = \{1, ..., K\}$ and $\mathbb{T} = \{1, ..., T\}$, respectively. Let $\mathbf{x}_t = (x_{1t}, ..., x_{Kt}) \in \mathbb{R}_+^K$ denote the observed quantity-vector at time $t \in \mathbb{T}$ with corresponding price-vector $\mathbf{p}_t = (p_{1t}, ..., p_{Kt}) \in \mathbb{R}_{++}^K$. It is assumed that budgets are exhaustive, so that budget shares can be expressed as:

$$w_{kt} = \frac{p_{kt} x_{kt}}{\mathbf{p}_t \mathbf{x}_t}, \tag{1}$$

for all $k \in \mathbb{K}$ and $t \in \mathbb{T}$. Consider the concept of revealed preferences: for two observations $s, t \in \mathbb{T}$, $\mathbf{x}_t$ is directly revealed preferred to $\mathbf{x}_s$ written $\mathbf{x}_t R^D \mathbf{x}_s$ if $\mathbf{p}_t \mathbf{x}_t \geq \mathbf{p}_t \mathbf{x}_s$, which means that $\mathbf{x}_t$ was chosen when in fact $\mathbf{x}_s$ also was affordable at prices $\mathbf{p}_t$. We say that $\mathbf{x}_t$ is revealed preferred to $\mathbf{x}_s$, written $\mathbf{x}_t R \mathbf{x}_s$, if there exists a sequence of observations $(t, u, v, ..., w, s) \in \mathbb{T}$ such that $\mathbf{x}_t R^D \mathbf{x}_u, \mathbf{x}_u R^D \mathbf{x}_v, ...,$ $\mathbf{x}_w R^D \mathbf{x}_s$. The data $\mathbb{D} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in \mathbb{T}}$ satisfies the Generalized Axiom of Revealed Preference (GARP) if $\mathbf{x}_t R \mathbf{x}_s$ implies $\mathbf{p}_s \mathbf{x}_s \leq \mathbf{p}_s \mathbf{x}_t$ (Varian, 1982).

The importance of GARP for empirical and theoretical work can be summarized in what has become known as Afriat's theorem (Afriat, 1967). We say that a utility function $U(\mathbf{x})$ rationalizes a data set $\mathbb{D} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t \in \mathbb{T}}$ if $\mathbf{p}_t \mathbf{x}_t \geq \mathbf{p}_t \mathbf{x}$ implies $U(\mathbf{x}_t) \geq U(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}_+^K$ and all $t \in \mathbb{T}$. Varian's (1982) formulation of Afriat's theorem states that there exist a continuous, strictly increasing and concave utility function rationalizing the data $\mathbb{D}$ if and only if $\mathbb{D}$ satisfies GARP. In practice, GARP can be efficiently implemented in standard statistical and mathematical software, and has become the most popular test-method in empirical applications of consumer rationality.[5]

[4] This index is called Algorithm 3 in Bronars (1987).

[5] See Demuynck and Hjertstrand (2019) for a recent survey of Afriat's theorem.

3 The partial uniform random power (PURP) index

The power of a revealed preference test, such as GARP, is defined as the probability of rejecting GARP, given that the consumption choices were generated from some type of irrational behavior. Bronars (1987) proposes a power index where the irrational behavior is based on Becker's (1962) uniformly random model. In his model, budget shares are uniformly distributed on the unit simplex, which implies that the consumption choices are uniformly distributed on the frontier of the budget set.

Bronars' uniformly random power index imposes the restriction that the *entire* expenditure is randomly allocated among the goods. However, it seem more empirically sensible to assume that a consumer only randomly allocates a *fraction* of her expenditure. In this model, the budget share, w_k^q, for any good $k \in \mathbb{K}$ corresponds to the weighted average between the observed budget share w_k and the uniformly random budget share, w_k^U, i.e.,

$$w_k^q = (1-\lambda)\, w_k + \lambda w_k^U. \tag{2}$$

The parameter $\lambda \in [0,1]$ can be interpreted as the fraction of expenditure that is randomly allocated.[6] This model encompasses a continuum of sub-models. At the one extreme $\lambda = 0$, it conform to when no part of the expenditure is being randomly allocated, in which case w_k^q is equal to the observed budget share, i.e., $w_k^q = w_k$ (See the third remark below). At the other extreme $\lambda = 1$, the model reduces to Bronars' pure model of uniformly random behavior, implying that w_k^q is equal to the uniformly random budget share, i.e., $w_k^q = w_k^U$. Thus, from footnote 6, we see that in Bronars' power index equal weight is put on all budget shares, i.e., $E\left[w_k^q\right] = E\left[w_k^U\right] = \frac{1}{K}$.

The model (2) is illustrated in the four graphs on the left-hand side of Figure 1. We assume $K = 3$ and that the observed budget shares are given by $(w_1, w_2, w_3) = (0.6, 0.3, 0.1)$ (marked with a red point). Each graph contain $5,000$ simulated points of (w_1^q, w_2^q, w_3^q) calculated from the model (2) for $\lambda = (0.25, 0.5, 0.75, 1.0)$.[7] We see that as λ increases, the area containing the observed shares grow and for $\lambda = 1$ (bottom left graph) it covers the entire unit simplex.

I propose to use (2) as the model of irrational behavior when calculating the power of GARP. The implementation of this power index, which I call the partial uniform random power (PURP) index, takes the following 8 steps:

1. *Set* $\lambda \in [0,1]$.

2. *Choose the number of simulations* S *and set* $m = 0$.

3. *For each* $t \in \mathbb{T}$*, generate the uniformly random budget shares* $\left\{w_{kt}^U\right\}_{k\in\mathbb{K}}$ *on the unit simplex from the Dirichlet distribution with all parameters set to one, i.e.,* $\left(w_{1t}^U, ..., w_{Kt}^U\right) \sim \mathrm{Dir}\,(1, ..., 1)$*. This can easily be done by first generating* $\mathrm{Gamma}(1,1)$ *random variables, denoted as* $\{G_{kt}\}_{k\in\mathbb{K}}$*, by drawing* K *independent uniformly random numbers* $u_{kt} \sim U_{(0,1)}$*, and calculate* $G_{kt} = -\ln u_{kt}$*. The Dirichlet random numbers are given by* $w_{kt}^U = G_{kt}/\sum_{j=1}^K G_{jt}$ for all $k \in \mathbb{K}$.

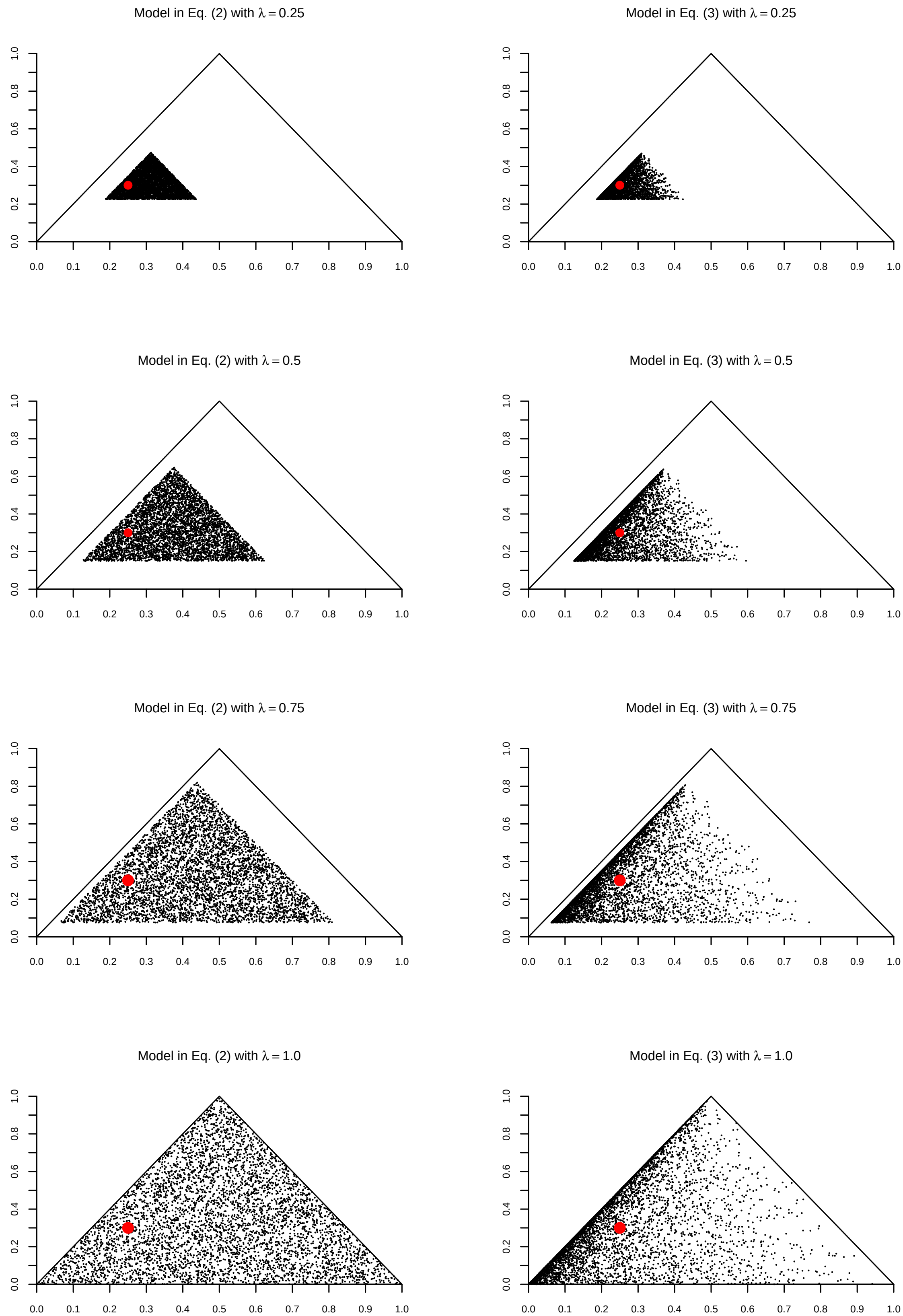
---

[6] Assuming that w_k^U are Dirichlet distributed random variables with parameters $(1, 1, ..., 1)$, the expected value, variance, covariance and correlation of the budget shares in the model (2) are:

$$E\left[w_k^q\right] = w_k + \lambda\left(\frac{1}{K} - w_k\right), \qquad \text{and} \qquad Var\left[w_k^q\right] = \lambda^2 \frac{(K-1)}{K^2\,(K+1)};$$
$$Cov\left[w_k^q, w_j^q\right] = -\lambda^2 \frac{1}{K^2\,(K+1)}, \qquad \text{and} \qquad Corr\left[w_k^q, w_j^q\right] = -\frac{1}{(K-1)} \quad \text{for} \quad k \neq j.$$

[7] Each simulated three-dimensional point represents a vector that is projected on the two-dimensional simplex in every graph. The vertices in the simplex are given by: $(1,0,0) \longrightarrow (0,0)$, $(0,1,0) \longrightarrow (1,0)$ and $(0,0,1) \longrightarrow (0.5,1)$.

Figure 1: 5,000 simulated points on the unit simplex from the model in Eq. (2) and the model in Eq. (3) for different values of λ. The red point are the observed budget shares $(w_1, w_2, w_3) = (0.6, 0.3, 0.1)$.



4. *For each* $t \in \mathbb{T}$, *calculate* $\{w^q_{kt}\}_{k\in\mathbb{K}}$ *from (2), where the observed budget shares* $\{w_{kt}\}_{k\in\mathbb{K}}$ *are calculated from (1).*

5. *Given prices,* p_{kt}, *and expenditure,* $\mathbf{p}_t\mathbf{x}_t$, *solve for the random quantities,* q_{kt}, *corresponding to* w^q_{kt} *as:*

$$q_{kt} = w^q_{kt}\frac{\mathbf{p}_t\mathbf{x}_t}{p_{kt}},$$

   and define $\mathbf{q}_t = (q_{1t}, ..., q_{Kt})$ *for all* $t \in \mathbb{T}$.

6. *If* $\{\mathbf{p}_t, \mathbf{q}_t\}_{t\in\mathbb{T}}$ *violates* GARP *then set* $m = m + 1$.

7. *Repeat steps 3-6* S *times.*

8. *The power is equal to* m/S.

Some remarks: First, the PURP index is just as easy to implement as Bronars' uniformly random power index, and can be applied using the same sorts of data used for revealed preference.

Second, the index can trace out the entire power curve against uniformly random expenditure allocation by implementing the model at each node in an equally-spaced grid for $\lambda \in [0, 1]$.

Third, the case $\lambda = 0$, in which no uniformly random data is added to the observed data, is special in the sense that it does not necessarily constitute a model of irrationality (i.e., it may be inconsistent with any alternative hypothesis). Indeed, this case has two interpretations depending on whether the observed data $\mathbb{D} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t\in\mathbb{T}}$ satisfy or violate GARP. On the one hand, if $\mathbb{D}$ *satisfy* GARP, then by Afriat's theorem, these data were generated under the null hypothesis of rationality. As such, implementing the 8-step procedure calculates the size (and not power) of the GARP test. Note that the size of any revealed preference test (e.g., GARP), by construction, is always zero.

On the other hand, if $\mathbb{D}$ *violate* GARP, then by Afriat's theorem, these data were generated under the alternative hypothesis of irrational behavior, and implementing the 8-step procedure calculates the power of the GARP test. In this case, the power is, by construction, always equal to 1.

Fourth, what is a reasonable value of λ? This is not pertained to the methods suggested here, but shared by other revealed preference procedures. In empirical revealed preference applications of "approximate utility maximization", for example, a cut-off point is chosen at which any cost-efficiency level below that point is considered as inconsistent with rational choice. In contrast, any point above the cut-off is considered sufficiently close to "exact" utility maximizing behavior and therefore deemed consistent with rationality. However, what consititutes a reasonable cut-off is subjective, and Varian (1990) argues that it should depend on the problem at hand, e.g., the number of observations, power of the test, and model under consideration.[8]

Since λ in the PURP index has a similar function as e in the theory of "approximate rationality", the same principle can be extended to λ. Accordingly, I argue that the choice of "cut-off" at which the fraction of random expenditure allocation is deemed reasonable should depend on the problem and data at hand. In particular, I do not take a direct stand on what constitutes a reasonable λ, but encourage readers to make up their own minds on what they think is a small or large fraction of random expenditure allocation in empirical applications of the PURP index.

Finally, following a suggestion of a reviewer, I provide an alternative interpretation of the PURP index. Suppose that the observed data $\mathbb{D} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t\in\mathbb{T}}$ satisfy GARP and consider the model (2). So

[8] A similar reasoning is applicable to standard statistical hypothesis testing, where the researcher is required to choose a nominal significance level, and one can argue that this choice should also depend on the problem at hand.

far, λ have been interpreted as the fraction of the expenditure that is randomly allocated. Instead, let us think of λ as measuring the degree of irrationality in the model (2), such that the model becomes "more irrational" for larger values of λ. That is, for large values of λ, the model (2) is "very irrational", while it is "less irrational" for lower values of λ. The case $\lambda = 0$ constitutes a perfectly rational model. Hence, in this framework, the PURP index for a fixed λ can be interpreted as the probability of rejecting GARP at the degree of irrationality λ, denoted Prob[GARP is rejected | λ]. Assuming that the derivative ∂Prob[GARP is rejected | λ] $/\partial\lambda$ is well-defined, this derivative can then (roughly) be interpreted as the "speed" at which the GARP test becomes rejectable at the degree of irrationality λ, given the observed data $\mathbb{D} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t\in\mathbb{T}}$. Similar to the PURP curve, it is possible to trace out the entire "speed"-curve by calculating the derivative ∂Prob[GARP is rejected | λ] $/\partial\lambda$ at each node in an equally-spaced grid for $\lambda \in [0,1]$ and then plotting the derivative against each node.

4 The centered partial uniform random power (cPURP) index

A possibly unappealing feature of the PURP index is that it does not generate budget shares that are centered on the observed budget shares.[9] In this section, I propose a simple modification of the PURP index that satisfies this property. As above, we consider a model where the budget shares, in this case denoted by w_k^z, are a linear combination of the observed budget share, w_k, and a random budget share denoted by w_k^C, i.e.,

$$w_k^z = (1-\lambda)\, w_k + \lambda w_k^C. \tag{3}$$

However, in contrast to the PURP index where the random budget shares follow a Dirichlet distribution with all parameters set to one, in this model, the random budget shares $\left(w_1^C, ..., w_K^C\right)$ follow a Dirichlet distribution with parameters $(K \times w_1, ..., K \times w_K)$. This ensures that the expected value of the random budget shares is equal to the observed shares, i.e., $E\left[w_k^C\right] = w_k$ for all $k \in \mathbb{K}$. Hence, the budget shares w_k^z in the model (3) are *always* centered on the observed shares (for any λ), i.e.,[10]

$$\begin{aligned} E\left[w_k^z\right] &= E\left[(1-\lambda)\, w_k + \lambda w_k^C\right] \\ &= (1-\lambda)\, w_k + \lambda E\left[w_k^C\right] \\ &= (1-\lambda)\, w_k + \lambda w_k \\ &= w_k. \end{aligned}$$

As before, the parameter $\lambda \in [0,1]$ is the fraction of expenditure that is randomly allocated. The graphs on the right-hand side of Figure 1 plots $5,000$ simulated shares of w_k^z from the model (3) for $\lambda = (0.25, 0.5, 0.75, 1.0)$. We see that the simulated shares for the model (3) are more scattered around the observed shares $(w_1, w_2, w_3) = (0.6, 0.3, 0.1)$ than in the model (2).

I propose the centered partial uniform random power (cPURP) index based on the model of irrationality in (3). The cPURP index is implemented as follows:

1-2. *Same as above.*

[9] The budget shares in the PURP index are centered on the observed budget shares only if $w_k = \frac{1}{K}$ for all $k \in \mathbb{K}$, which implies $E\left[w_k^q\right] = w_k = \frac{1}{K}$.

[10] The variance, covariance and correlation of the budget shares in the model (3) are:

$$Var\left[w_k^z\right] = \lambda^2 \frac{w_k\,(1-w_k)}{(K+1)}, \quad Cov\left[w_k^z, w_j^z\right] = -\lambda^2 \frac{w_k w_j}{(K+1)} \quad \text{and} \quad Corr\left[w_k^z, w_j^z\right] = -\sqrt{\frac{w_k w_j}{(1-w_k)\,(1-w_j)}} \quad \text{for } k \neq j.$$

3. *For each* $t \in \mathbb{T}$, *generate the random budget shares* $\left\{w_{kt}^C\right\}_{k\in\mathbb{K}}$ *on the unit simplex from the Dirichlet distribution* $\left(w_{1t}^C, ..., w_{Kt}^C\right) \sim \text{Dir}\left(K \times w_{1t}, ..., K \times w_{Kt}\right)$. *These variables can be simulated by first drawing* $G_{kt} \sim \text{Gamma}(K \times w_{kt}, 1)$ *and then setting* $w_{kt}^C = G_{kt} / \sum_{j=1}^{K} G_{jt}$ *for all* $k \in \mathbb{K}$.

4. *For each* $t \in \mathbb{T}$, *calculate* $\{w_{kt}^z\}_{k\in\mathbb{K}}$ *from (3), where the observed budget shares* $\{w_{kt}\}_{k\in\mathbb{K}}$ *are calculated from (1).*

5. *Given prices,* p_{kt}, *and expenditure,* $\mathbf{p}_t\mathbf{x}_t$, *solve for the random quantities,* z_{kt}, *corresponding to* w_{kt}^z *as:*

$$z_{kt} = w_{kt}^z \frac{\mathbf{p}_t\mathbf{x}_t}{p_{kt}},$$

and define $\mathbf{z}_t = (z_{1t}, ..., z_{Kt})$ *for all* $t \in \mathbb{T}$.

6. *If* $\{\mathbf{p}_t, \mathbf{z}_t\}_{t\in\mathbb{T}}$ *violates* GARP *then set* $m = m + 1$.

7-8. *Same as above.*

Thus, the cPURP index differs from the PURP index only by step 3. The cPURP index reduces to the PURP index when the observed budget shares $(w_{1t}, ..., w_{Kt})$ are equal across goods, i.e., if $w_{kt} = \frac{1}{K}$ for all $k \in \mathbb{K}$ and $t \in \mathbb{T}$, in which case, we have $\text{Dir}\left(K \times w_{1t}, ..., K \times w_{Kt}\right) \overset{d}{=} \text{Dir}\left(1, ..., 1\right)$.

5 Applications

In this section, I apply the PURP and cPURP indices to experimental data collected by Andreoni and Miller (2002) and Fisman, Kariv and Markovits (2007) to study if altruistic choices of experimental subjects are rational.

5.1 Data from Andreoni and Miller (2002)

Andreoni and Miller (2002, AM) employed a generalized dictator game in which one subject (the dictator) allocates token endowments between himself and an anonymous other subject with different transfer rates. The payoffs of the dictator and the beneficiary are interpreted as two distinct goods and the transfer rates as the price ratio. The experiment was split in two parts, where 142 subjects (Group 1) faced $T = 8$ decision rounds while 34 subjects (Group 2) faced $T = 11$ decision rounds. AM used GARP to test for rationality and found that 13 subjects (9%) in Group 1 and 5 subjects (15%) in Group 2 violated GARP. They also reported a Bronars' uniformly random power index of 78.1% for Group 1 and 94.7% for Group 2.[11] Andreoni et al. (2013) provided a more detailed power analysis of these data by applying their own three indices and all three of Bronars' power indices. I complement the analyses in AM and Andreoni et al. (2013) by calculating the PURP and cPURP indices for those subjects in Groups 1 and 2 that satisfy GARP. I implement the indices at each node using a grid of 0.02 for λ (starting at 0 and ending at 1) and the number of simulations in step 2 is set to $S = 10,000$.

Figure 2(a) traces out the power curves for the 129 subjects in Group 1 that satisfy GARP using the PURP index. The left panel in Table 1 gives summary statistics over these 129 subjects for different values of λ. The power curves for most subjects are essentially zero up to a random expenditure allocation of 20% ($\lambda = 0.2$). For a random expenditure allocation of 50% ($\lambda = 0.5$), we see that 75% (Q3) of all

[11] Because the subjects in each group faced the same prices and budgets, Bronars' uniformly random power index is the same for every subject in each group.

Figure 2: Power plots for the PURP and cPURP indices over all subjects that satisfy GARP in Groups 1 and 2.

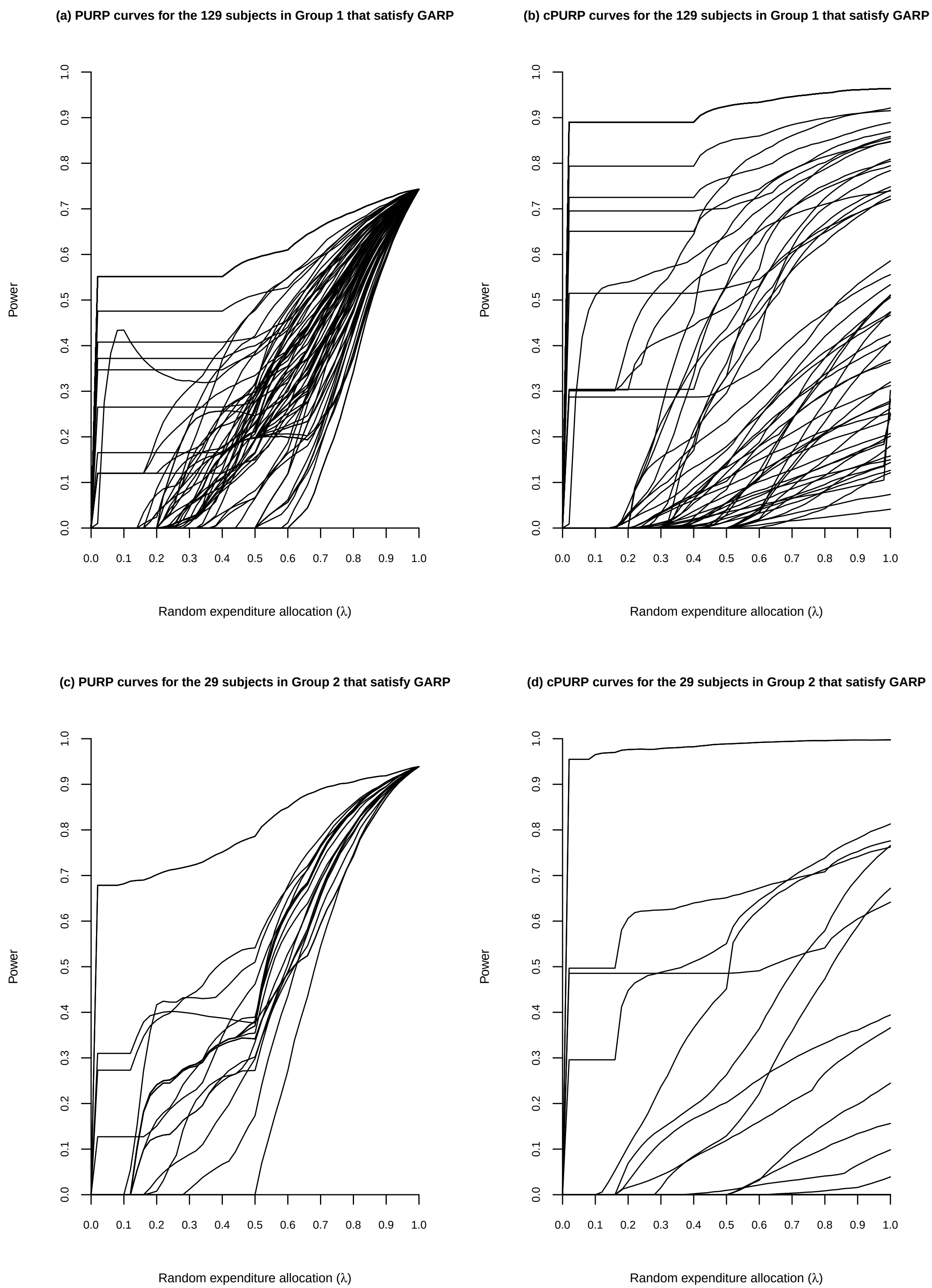


Table 1: Summary statistics for the PURP and cPURP indices over the 129 subjects in Group 1 that satisfy GARP for different values of λ.

λ	Mean	Min	Q1	Med.	Q3	Max	Mean	Min	Q1	Med.	Q3	Max
	PURP						cPURP					
0.1	0.121	0	0	0	0.120	0.551	0.201	0	0	0	0.301	0.890
0.2	0.124	0	0	0	0.132	0.551	0.203	0	0	0	0.303	0.890
0.3	0.163	0	0.055	0.055	0.226	0.551	0.217	0	0	0	0.331	0.890
0.4	0.225	0	0.159	0.159	0.274	0.551	0.234	0	0	0	0.451	0.890
0.5	0.279	0	0.197	0.212	0.346	0.592	0.261	0	0	0.022	0.530	0.925
0.6	0.371	0	0.272	0.386	0.454	0.610	0.285	0	0	0.044	0.630	0.934
0.7	0.467	0.120	0.393	0.489	0.547	0.661	0.313	0	0	0.098	0.702	0.946
0.8	0.572	0.343	0.528	0.583	0.631	0.693	0.334	0	0	0.136	0.746	0.954
0.9	0.675	0.589	0.658	0.681	0.696	0.723	0.353	0	0	0.163	0.777	0.961
1	0.743	0.743	0.743	0.743	0.743	0.743	0.370	0	0	0.239	0.806	0.963

Notes: Q1: 1st quartile (25th percentile), Med.: Median (2nd quartile, 50th percentile) and Q3: 3rd quartile (75th percentile).

subjects have a power of 34.6% or less. The average power over all subjects is 46.7% for a random expenditure allocation of 70% ($\lambda = 0.7$) and increases to 50% at $\lambda \simeq 0.72$. In general, the power increases more rapidly for larger values of λ. For example, while the average power increases by 15.8% from $\lambda = 0.1$ to $\lambda = 0.5$, it increases by 46.4% from $\lambda = 0.5$ to $\lambda = 1$. However, this is not the case for the cPURP index, which is generally higher than the PURP index for lower values of λ (≤ 0.4), but cPURP is lower than PURP for higher values of λ (> 0.4). A likely explanation for this is that a few subjects have a high power as measured by cPURP at low values of λ (See Q3- and Max-columns). This can be better seen by comparing Figures 2(a) and (b), where the cPURP-curves in Figure 2(b) are more spread out and heterogeneous than the PURP-curves. In other words, there is a larger difference in power between subjects when measured by the cPURP index. This is also visable from the right panel in Table 1, which gives summary statistics over the 129 subjects for different values of λ in the cPURP index. Specifically, half of the subjects have a power of at most 2.2% for a random expenditure allocation of 50%, and the average power is 37% for a random expenditure allocation of 100%. Further illustrating the heterogeneity over subjects in the cPURP index, note from the last column in Table 1 that the power for a few subjects is close to 1 for values of λ near zero.

Figure 2(c) traces out the power curves for the 29 subjects in Group 2 that satisfy GARP using the PURP index. The left panel in Table 2 gives summary statistics over the 29 subjects for different values of λ. As expected from the larger number of decision rounds, the power is generally higher for subjects in this group than in Group 1.

Next, consider the results for the cPURP index in Figure 2(d) and the right panel in Table 2. The cPURP index is considerably lower than the PURP index for these subjects. The power for half of the subjects is equal to zero even at a random allocation of 100%. Moreover, contrary to what one might expect (given the larger number of decision rounds), the cPURP index is generally lower for subjects in Group 2 than in Group 1. A possible explanation for this is that the cPURP index does not only depend on the observed budget shares though the choice of λ, but also depend on the shares in the generation

Table 2: Summary statistics for the PURP and cPURP indices over the 29 subjects in Group 2 that satisfy GARP for different values of λ.

λ	Mean	Min	Q1	Med.	Q3	Max	Mean	Min	Q1	Med.	Q3	Max
	PURP						cPURP					
0.1	0.071	0	0	0	0	0.682	0.110	0	0	0	0	0.965
0.2	0.243	0	0.160	0.241	0.241	0.702	0.128	0	0	0	0.039	0.976
0.3	0.287	0	0.250	0.282	0.282	0.721	0.142	0	0	0	0.122	0.979
0.4	0.341	0	0.329	0.334	0.337	0.751	0.155	0	0	0	0.174	0.982
0.5	0.385	0	0.342	0.380	0.380	0.786	0.167	0	0	0	0.217	0.989
0.6	0.587	0.272	0.488	0.623	0.626	0.850	0.190	0	0	0	0.281	0.992
0.7	0.717	0.545	0.664	0.746	0.746	0.889	0.211	0	0	0	0.390	0.994
0.8	0.824	0.742	0.803	0.841	0.845	0.905	0.229	0	0	0	0.490	0.996
0.9	0.896	0.870	0.887	0.903	0.905	0.919	0.250	0	0	0	0.593	0.997
1	0.939	0.939	0.939	0.939	0.939	0.939	0.266	0	0	0	0.649	0.998

of the Dirichlet random numbers (in step 3).[12] Hence, for the cPURP index, increasing the number of observations matter less if one wishes to obtain a more powerful test.

5.2 Data from Fisman, Kariv and Markovits (2007)

Like AM, Fisman, Kariv, and Markovits (2007, FKM) conducted a laboratory experiment to study individual preferences for giving. They also employ a generalized dictator game, where the payoffs of the dictator and the recipient can be interpreted as goods. While the experiment in AM consisted of 8 and 11 decision rounds, the experiment in FKM consisted of 50 decision rounds with 76 subjects. Thus, because of the many more decision rounds in the FKM-experiment, the power of the revealed tests used to check if subjects' choices are rational should have a higher discrimatory power against irrational behavior than in AM. Heufer and Hjertstrand (2017) provide a detailed analysis of the power for these data, but only considered Bronars (1987) alternative hypothesis of pure uniformly random behavior.

Most subjects in the FKM data violate GARP, and to test if the subjects' choices obey the notion of "approximate utility maximization" (as briefly discussed in Section 3), FKM calculate the minimal adjustment of expenditure required to make the data consistent with GARP, given by the Afriat efficiency index (AEI).[13] Since adjusting expenditure by the AEI gives a less stringent test, Heufer and Hjertstrand (2017) analyze the loss in power for data that are adjusted by the AEI. To calculate the power, they employ the following three-step procedure: (i) compute the AEI using the observed data;

[12] Note that this is the difference between cPURP and PURP since the generation of the uniformly random budget shares in the latter is fixed (drawn from a Dirichlet distribution with all parameters set to one).

[13] GARP can be straightforwardly modified to test for "approximate utility maximization" and calculate the AEI. Consider an efficiency level $0 \leq e \leq 1$: $\mathbf{x}_t$ is directly revealed preferred to $\mathbf{x}_s$ at efficiency level e, written $\mathbf{x}_t R_e^D \mathbf{x}_s$ if $e\mathbf{p}_t\mathbf{x}_t \geq \mathbf{p}_t\mathbf{x}_s$. The relation R_e is the transitive closure of R_e^D and defined analogously as R. The scalar e can be thought of as how much less the potential expenditure on a bundle $\mathbf{x}_t$ has to be before we will consider it worse than the observed choice $\mathbf{x}_s$. For example, if e is 0.9, we will only count bundles whose cost is less than 90% of an observed choice as being revealed worse than that choice. The data $\mathbb{D} = \{\mathbf{p}_t, \mathbf{x}_t\}_{t\in\mathbb{T}}$ satisfies the Generalized Axiom of Revealed Preference at efficiency level e (GARP$_e$) if $\mathbf{x}_t R_e \mathbf{x}_s$ implies $e\mathbf{p}_s\mathbf{x}_s \leq \mathbf{p}_s\mathbf{x}_t$. The AEI is defined as the maximal e such that the data obey GARP$_e$.

Table 3: Summary statistics for the PURP and cPURP indices for GARP over subjects in the FKM data for different values of λ.

λ	Mean	Min	Q1	Med.	Q3	Max	Mean	Min	Q1	Med.	Q3	Max
	PURP						cPURP					
0.1	0.411	0	0.126	0.361	0.716	0.985	0.473	0	0.175	0.505	0.701	0.994
0.2	0.524	0	0.165	0.426	0.960	1.000	0.537	0	0.282	0.560	0.818	1.000
0.3	0.621	0	0.266	0.755	0.992	1.000	0.603	0	0.380	0.634	0.903	1.000
0.4	0.722	0	0.469	0.914	0.999	1.000	0.659	0	0.469	0.756	0.957	1.000
0.5	0.800	0.007	0.723	0.986	1.000	1.000	0.691	0	0.514	0.808	0.973	1.000
0.6	0.865	0.016	0.892	0.999	1.000	1.000	0.718	0	0.546	0.882	0.984	1.000
0.7	0.909	0.032	0.966	1.000	1.000	1.000	0.739	0	0.556	0.919	0.996	1.000
0.8	0.944	0.054	0.992	1.000	1.000	1.000	0.756	0	0.557	0.945	0.999	1.000
0.9	0.967	0.171	0.999	1.000	1.000	1.000	0.768	0	0.578	0.969	1.000	1.000
1	0.985	0.442	1.000	1.000	1.000	1.000	0.778	0	0.603	0.981	1.000	1.000

then, (ii) generate random datasets using Bronars' approach; and, finally, (iii) calculate the fraction of sets violating revealed preference where expenditures are adjusted by the AEI (computed in the first step). In addition to applying GARP, Heufer and Hjertstrand (2017) also test for "approximate homothetic utility maximization" by applying the homothetic axiom of revealed preference (HARP). See Heufer and Hjertstrand (2017) for a detailed discussion of HARP and for the definition of "approximate homothetic utility maximization", and also for the definition of the homothetic efficiency index (HEI), which is analogous to the AEI in terms of homotetic utility maximization.

In this application, I apply the PURP and cPURP indices to the FKM-data using the three-step procedure. This analysis gives a more complete picture of the power and provides at least a partial answer to whether it is meaningful to infer if these data were generated by (homothetically) rational subjects even when expenditures are deflated by the AEI/HEI. I implement the indices at each node using a grid of 0.02 for λ (starting at 0 and ending at 1) and the number of simulations in step 2 is set to $S = 1,000$.

Table 3 gives the results for GARP when expenditures are deflated by the AEI (the last column replicates the results in Heufer and Hjertstrand, 2017). In general, the power is higher at lower levels of random expenditure allocation than in AM's data, which is most likely due to the many more decision rounds in the FKM-data. This is even more evident for the cPURP index, which is considerably higher in the FKM-data. As in AM's data, the cPURP index for most subjects is higher than the PURP index at lower levels of expenditure allocations but, again, lower at higher levels of λ.

Next, we turn to the results for HARP (recall that expenditures are deflated by the HEI), which is given in Table 4 (the last column replicates the results in Heufer and Hjertstrand, 2017). Note that since the HEI cannot be higher than the AEI, the HARP test may actually be less stringent than the GARP test (when deflated by the HEI and AEI, respectively). Hence, with expenditures deflated by the HEI and AEI, the power may be lower for HARP than for GARP.

The PURP results in the left panel of Table 4 are very similar to the PURP results in Table 3. The cPURP results on the right panel of Table 4 are uniformly higher than the cPURP results in Table 3. If the HEI for the subjects is considered large enough and therefore sufficiently close to "exact" homothetic utility maximizing behavior, then the results in Table 4 may be taken as some evidence that

Table 4: Summary statistics for the PURP and cPURP indices for HARP over subjects in the FKM data for different values of λ.

λ	Mean	Min	Q1	Med.	Q3	Max	Mean	Min	Q1	Med.	Q3	Max
	PURP						cPURP					
0.1	0.408	0.005	0.081	0.184	0.928	1.000	0.545	0	0.423	0.555	0.711	1.000
0.2	0.515	0.005	0.119	0.372	1.000	1.000	0.622	0	0.477	0.656	0.835	1.000
0.3	0.617	0.006	0.205	0.648	1.000	1.000	0.688	0	0.545	0.744	0.914	1.000
0.4	0.708	0.009	0.403	0.947	1.000	1.000	0.733	0	0.585	0.836	0.962	1.000
0.5	0.793	0.010	0.680	0.990	1.000	1.000	0.769	0	0.619	0.903	0.993	1.000
0.6	0.863	0.013	0.888	1.000	1.000	1.000	0.794	0	0.659	0.961	0.999	1.000
0.7	0.915	0.020	0.983	1.000	1.000	1.000	0.808	0	0.711	0.987	1.000	1.000
0.8	0.954	0.057	0.999	1.000	1.000	1.000	0.817	0	0.738	0.992	1.000	1.000
0.9	0.974	0.194	1.000	1.000	1.000	1.000	0.824	0	0.755	0.997	1.000	1.000
1	0.989	0.626	1.000	1.000	1.000	1.000	0.829	0	0.773	0.999	1.000	1.000

the "approximate" homothetic utility maximization model describes the behavior of the subjects better than the "approximate" standard utility maximization model. Heufer and Hjertstrand (2018) propose to use the "approximate" homothetic utility maximization model to recover preferences by calculating bounds on indifference curves in experimental data.[14] One may therefore argue that the results in Table 4 gives further motivation for using Heufer and Hjertstrand's (2018) methods to recover preferences (as opposed to using the "approximate" standard utility maximization model).

6 Conclusions

This paper has proposed new power indices for revealed preference tests. The indices are based on models of irrationality where the consumer only randomly allocates a certain fraction of expenditure. This is different from other power indices, where the consumer is assumed to randomly allocate the entire expenditure. A novel feature of the new power indices is that they can be used to trace out the entire power curve against random expenditure allocation. The indices are easy to implement and can be readily combined with other power indices such as Heufer's (2014) conditional power index.

[14] Imposing homotheticity on preferences have some obvious advantages since, under homotheticity, everything there is to know about an agent's preference is implicit in a single indifference set. In particular, when there are only two goods, which is common in experimental economics, a preference can therefore be completely summarized graphically by showing a single indifference curve.

References

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

Afriat, S.N. (1972). Efficiency estimation of production functions. *International Economic Review* **13**, p.568–598.

Andreoni, J., B.J. Gillen and W.T. Harbaugh (2013). The power of revealed preference tests: Ex-Post evaluation of experimental design. Manuscript, Univ. California, San Diego.

Andreoni, J. and J. Miller (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* **70**, p.737-750.

Becker, G.S. (1962). Irrational behavior and economic theory. *Journal of Political Economy* **70**, p.1–13.

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

Dean, M. and D. Martin (2016). Measuring rationality with the minimum cost of revealed preference violations. *Review of Economics and Statistics* **98**, p.524-534.

Demuynck, T. (2020). Markov chain monte carlo procedure to generate revealed preference consistent datasets. Working Papers ECARES 2020-33, ULB – Universite Libre de Bruxelles.

Demuynck T. and P. Hjertstrand (2019). Samuelson's approach to revealed preference theory: Some recent advances. In: Cord R., Anderson R., Barnett W. (eds) *Paul Samuelson Master of modern economics. Remaking Economics: Eminent Post-War Economists.* Palgrave Macmillan, London.

Echenique, F., S. Lee, and M. Shum (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy* **119**, p.1201-1223.

Fisman, R., S. Kariv and D. Markovits (2007). Individual preferences for giving. *American Economic Review* **97**, p.1858–1876.

Heufer, J. (2014). Generating random optimising choices. *Computational Economics* **44**, p.295-305

Heufer, J. and P. Hjertstrand (2017). Homothetic efficiency: Theory and applications. *Journal of Business and Economic Statistics* **37**, p.235-247.

Heufer, J. and P. Hjertstrand (2018). Homothetic preferences revealed. *Journal of Economic Behavior and Organization* **157**, p.602-614.

Houtman, M. and J.A.H. Maks (1985). Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve Methoden* **89**, p.89-104.

Jones, B.E. and D.L. Edgerton (2009). Testing utility maximization with measurement errors in the data. *Advances in Econometrics* **24**, *Measurement error: Consequences, Applications and Solutions*, 199-236.

Russell, R. R. Breunig and C.-H. Chiu (1998). Aggregation and econometric analysis of supply and demand. In *Handbook of Applied Economic Statistics*, D. Giles and A. Ullah (Eds.), Marcel Dekker, New York, p.177–235.

Varian, H.R. (1982). The nonparametric approach to demand analysis. *Econometrica* **50**, p.945–974.

Varian, H.R. (1990). Goodness-of-fit in optimizing models. *Journal of Econometrics* **46**, p.125-140.

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

Varian, H.R. (2006). Revealed preference. In *Samuelsonian Economics and the 21st Century*, M. Szenberg, L. Ramrattand, and A.A. Gottesman (Eds.), Oxford University Press, Oxford, p.99-115.