THE UNIVERSITY OF WARWICK

January 2012 No.72

A Three-Stage Experimental Test of Revealed Preference

Peter Hammond & Stefan Traub
Department of Economics & CAGE, University of Warwick
Department of Economics & ZeS, University of Bremen

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE Competitive Advantage in the Global Economy

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

A Three-Stage Experimental Test of Revealed Preference

Peter Hammond[a,*], Stefan Traub[b]

[a]*Department of Economics & CAGE, University of Warwick, Coventry, UK*
[b]*Department of Economics & ZeS, University of Bremen, Germany*

29th January, 2012

Abstract
A powerful test of Varian's (1982) generalised axiom of revealed preference (GARP) with two goods requires the consumer's budget line to pass through two demand vectors revealed as chosen given other budget sets. In an experiment using this idea, each of 41 student subjects faced a series of 16 successive grouped portfolio selection problems. Each group of selection problems had up to three stages, where later budget sets depended on that subject's choices at earlier stages in the same group. Only 49% of subjects' choices were observed to satisfy GARP exactly, even by our relatively generous non-parametric test.

JEL classification: C91, D83
Keywords: Rationality, revealed preference, uncertainty

*Corresponding author. Department of Economics, University of Warwick, Coventry CV4 7AL, UK; phone +44 (0)24 765 23052, fax +44 (0)24 765 23032, e-mail p.j.hammond@warwick.ac.uk.

1 Introduction

1.1 Non-Parametric Tests of GARP

Varian (1982) in particular has emphasised how easily even a rational consumer could exhibit demand behaviour that fails rationality tests based on estimating preference parameters. As an alternative, Varian proposed more robust non-parametric tests of Samuelson's (1938) revealed preference theory that are based on Afriat's (1973) theoretically derived inequalities. This approach seems ideally suited to controlled laboratory experiments, where the price and income changes needed to test the axioms are easy to implement, and changes of taste can largely be ruled out. Also, in general violations of revealed preference could perhaps be explained by errors in observation, but hardly in experimental settings. Accordingly, several papers have followed Sippel's (1997) pioneering application of non-parametric tests to experimental data. Depending on the experimental design, however, including the population of experimental subjects and the test method, past experimental studies have produced estimates of the proportion of subjects whose demands satisfy GARP which range widely from below 10% to almost 100%.

Such results raise the fundamental question whether or not to allow for decision errors when testing revealed preference theory. On the one hand, *normative* decision theory does not condone even the slightest inconsistency; such a strict test has enormous statistical power but an impractically small size. On the other hand, allowing for random decision errors significantly reduces the power of a test in discriminating between rational and random behaviour.

Following Varian's (1982) own suggestion, Sippel (1997) and most successors have based their tests on Afriat's (1973) efficiency index. Suppose a consumer has been observed choosing the bundle $\mathbf{x}^1$ when the price vector was $\mathbf{p}^1$. By definition $\mathbf{x}^1$ is revealed preferred to any alternative bundle $\mathbf{x}^2$ satisfying $\mathbf{p}^1\mathbf{x}^2 < \mathbf{p}^1\mathbf{x}^1$. Suppose nevertheless that the same consumer were also observed choosing the bundle $\mathbf{x}^2$ when the price vector is $\mathbf{p}^2$, where $\mathbf{p}^2\mathbf{x}^2 > \mathbf{p}^2\mathbf{x}^1$. This would imply that $\mathbf{x}^2$ is revealed preferred to $\mathbf{x}^1$, and so violate GARP. The *Afriat efficiency index* of the choice $\mathbf{x}^2$ is the ratio $\mathbf{p}^1\mathbf{x}^2/\mathbf{p}^1\mathbf{x}^1$, which is evidently less than 1.

Allowing choices whose Afriat efficiency index is less than one relaxes the GARP axiom, and so increases considerably the corresponding measure of how well subjects' choices comply with GARP. This increase in measured rationality, however, comes with a dramatic decrease in statistical power.

For example, consider a budget of \$100, along with two budget lines determined by the respective price vectors $\mathbf{p}^1 = (1.25, 1)$ and $\mathbf{p}^2 = (1, 1.25)$, as illustrated in Fig. 1. Assume too that at prices $\mathbf{p}^1$ a person chooses the consumption bundle $\mathbf{x}^1 = (x^1_A, x^1_B) = (64, 20)$, or indeed any other bundle on the line segment joining the end point Q to the point $P = (44\frac{4}{9}, 44\frac{4}{9}) \approx (44.4, 44.4)$ where the two budget lines intersect. Then it is straightforward to show that at prices $\mathbf{p}^2$ the *supporting set* of consumption bundles satisfying GARP consists of the line segment joining P to the end point $Q' = (100, 0)$. Assuming a uniform distribution of choices along this second budget line, there is a probability of $\frac{5}{9} \approx 55.6\%$ that a player who chooses at random will satisfy GARP.

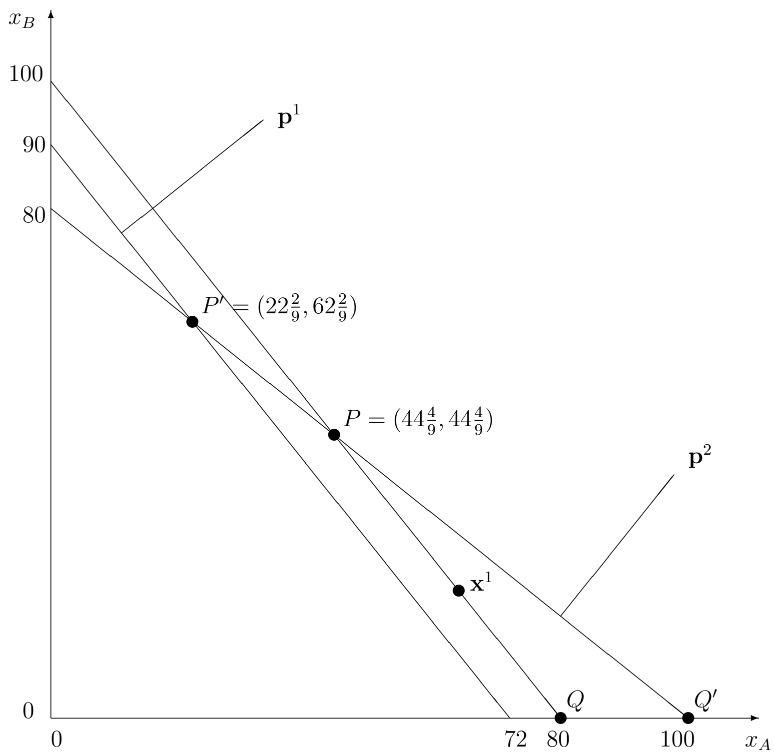

Figure 1: Basic Example

Allowing an Afriat efficiency index of 0.9, however, which is equivalent to throwing away \$10 at prices $\mathbf{p}^1$, moves the intersection of the two budget lines down to the point $P' = (22\frac{2}{9}, 62\frac{2}{9}) \approx (22.2, 62.2)$. This extends the supporting set to the line segment $P'Q'$, so the chance of a random choice being classified as rational rises to $\frac{7}{9} \approx 77.8\%$.

We refer the interested reader to their paper to Andreoni and Harbaugh (2008) for an extensive recent discussion of the pros and cons of several different power indices for revealed preference tests, including that of Bronars (1987). Instead, we now proceed directly to the experimental design involved in our own more direct test.

1.2 A Three-Stage Direct Test

Consider any list $s^n = (\mathbf{p}^i, \mathbf{x}^i)_{i=1}^n$ of n pairs of successive price and quantity vectors that satisfy both GARP and, for each $i = 1, \ldots, n$, the normalization $\mathbf{p}^i\mathbf{x}^i = 1$. Let $\mathbf{p}^{n+1}$ be any previously unobserved price vector. Then Varian (1982, 2006) defines the *supporting set* $S(\mathbf{p}^{n+1}; s^n)$ of consumption bundles $\mathbf{x}^{n+1}$ as those for which the extended sequence $(\mathbf{p}^i, \mathbf{x}^i)_{i=1}^{n+1}$ also satisfies both GARP and, for each $i = 1, \ldots, n+1$, the normalization $\mathbf{p}^i\mathbf{x}^i = 1$. As Varian (1982) notes, the supporting set describes "what choice a consumer will make if his choice is to be consistent with the preferences revealed by his previous behavior" (p. 957).

Our new experimental design uses Varian's (1982) supporting set directly. Moreover, unlike previous tests of GARP, we seek to increase the power of our tests by adjusting later budget lines to the consumer's earlier choices. Indeed, when teaching intermediate microeconomics, it is usual to explain the revealed preference axiom in a two-stage process. First it is assumed that a consumer chooses a (two-dimensional) commodity bundle $\mathbf{x}^1$ at the price vector $\mathbf{p}^1$. Second, one considers the consumer's demands when faced with a new price vector $\mathbf{p}^2$ and a new budget line $\mathbf{p}^2\mathbf{x} = \mathbf{p}^2\mathbf{x}^1$ that passes through the originally chosen bundle $\mathbf{x}^1$. The usual revealed preference axiom, of course, implies that the new bundle $\mathbf{x}^2$ should satisfy $\mathbf{p}^1\mathbf{x}^2 > \mathbf{p}^1\mathbf{x}^1$.

Our experiment considers an obvious three-stage extension. The first two stages involve observing the consumer choosing the two bundles $\mathbf{x}^1$ and $\mathbf{x}^2$ at the respective price vectors $\mathbf{p}^1$ and $\mathbf{p}^2$. Revealed preference requires the chosen bundles to satisfy both $\mathbf{p}^1\mathbf{x}^2 > \mathbf{p}^1\mathbf{x}^1$ and $\mathbf{p}^2\mathbf{x}^1 > \mathbf{p}^2\mathbf{x}^2$. Provided this condition was satisfied, subjects faced a third stage that involves a new price vector $\mathbf{p}^3$ satisfying $\mathbf{p}^3\mathbf{x}^1 = \mathbf{p}^3\mathbf{x}^2$. In the two commodity case we consider, this determines the third stage budget line $\mathbf{p}^3\mathbf{x} = \mathbf{p}^3\mathbf{x}^1$ uniquely. Revealed preference will be satisfied provided that the consumer's third-stage choice $\mathbf{x}^3$ is on the segment of the third budget line between the first two choices $\mathbf{x}^1$ and $\mathbf{x}^2$.

In our experiment, subjects were actually confronted with a series of 16 grouped portfolio-selection problems, each group involving up to three stages like this. As in the most important precursor to our own work, Choi *et al.* (2007b), we study a portfolio selection problem for two reasons. First, these authors were the first to test revealed preference theory with data from risky decision making. Indeed, as far as we are aware, other scientists have yet to reproduce their results. Second, their graphical interface seemed highly appropriate and was relatively easy to adapt. Third, they reported particularly high consistency rates among their subjects, which suggests that applying a more powerful test could be fruitful.

The paper is organised as follows. The next section 2 describes our experiment in more detail. Then Section 3 explains our nonparametric test procedure. The results are presented in Section 4. Section 5 concludes.

2 Details of the Experiment

2.1 Typical Decision Problem

As in Choi *et al.* (2007a, b), in each of our decision problems there were two states of nature $s = \{A, B\}$ and two associated Arrow securities, each yielding a payoff of one "token" of experimental currency in one state and nothing in the other. Following the usual random lottery incentive system, at the end of the experiment one decision problem was selected at random and each token won in that decision problem was converted into £0.20 of UK currency. In each decision problem, subjects had to split an initial endowment of 100 tokens between the two Arrow securities. In principle, their choices had to satisfy the budget constraint $p_A x_A + p_B x_B = 100$, where p_s denotes the price and x_s the demand for Arrow security s. In practice, in order to represent the allocation problem sensibly on the computer screen, prices were rounded off to the first decimal place, and subjects could only choose nonnegative integer amounts of each security. In addition to the budget constraint $p_A x_A + p_B x_B \leq 100$, subjects were restricted to pairs (x_A, x_B) of nonnegative integers immediately below the budget line. Specifically, we allowed any nonnegative

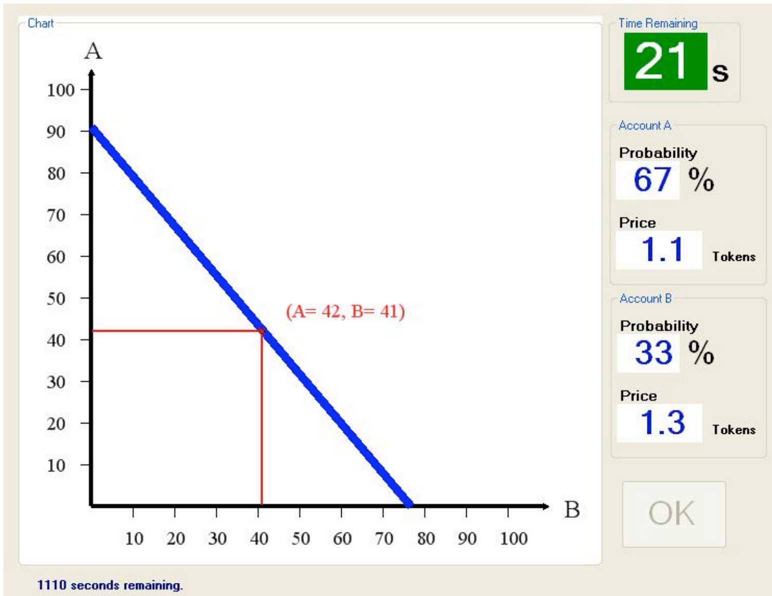

Figure 2: Example screen

integer allocation satisfying

$$100 - \max\{p_A, p_B\} < p_A x_A + p_B x_B \leq 100.$$

Figure 2 reproduces an example of what an experimental subject could see on the computer screen when faced with any of the choice problems. As soon as a new decision problem appeared, the mouse pointer became visible at its default position in the upper right-hand corner of the screen. When the mouse pointer was close enough to the nearest feasible allocation, that allocation was indicated by two numbers and by associated reference lines marked in red. This information remained visible until the mouse pointer had been moved far enough away from this allocation. If applicable, the next allocation was then displayed.

Subjects could also "fix" and later "release" an allocation by clicking the left mouse button. Once a portfolio was fixed, then even if the mouse pointer was moved, the numbers and reference lines turned green and stayed visible on the screen until they were released. To choose this indicated portfolio and proceed to the next decision problem, a subject could simply click the OK button near the lower right-hand corner of the screen.

Some slight time pressure was introduced in order to impose a "cost" of collecting information. The upper right-hand corner of the screen therefore displayed how many seconds remained out of the original 30 allocated for each choice. When time ran out, if the mouse pointer was over a feasible allocation, or if one had been fixed by an earlier mouse click, then that portfolio was recorded as the subject's final choice. Otherwise a missing value was recorded for that choice problem. In fact, no subject in our experiment ever exceeded the time limit.

Figure 3 illustrates the basic experimental setup for a scenario where $p_A = 1.5$, $p_B = 1$, and the probability of state A is $\pi = 0.5$. The solid line represents the budget constraint with slope $-p_B/p_A = -1.5$. The dashed $45°$-line marks all portfolios for which $x_A = x_B$. It intersects the budget line at the indicated *safe portfolio* ($x_A = x_B = 40$).

The second dashed line is the graph of the expected value

$$EV(x_B) = \pi x_A + (1 - \pi)x_B = \frac{\pi}{p_A}(100 - p_B x_B) + (1 - \pi)x_B$$

of each portfolio as a function of x_B alone, as one moves along the budget line. In figure 3 its slope is the positive fraction $1/6$. Hence, portfolios to the left of the safe portfolio are stochastically dominated.

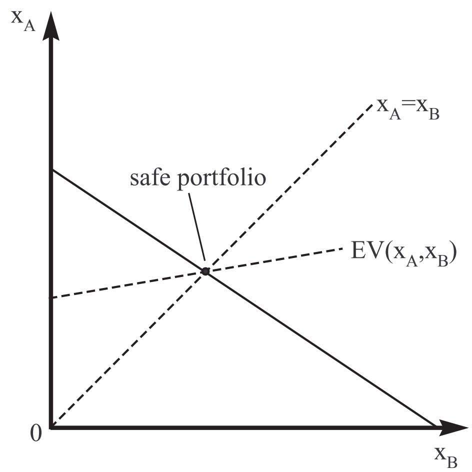

Figure 3: A first-stage choice problem with $p_A = 1.5$, $p_B = 1$, $\pi = 0.5$

2.2 First Stage

Each subject in the experiment faced 16 rounds of successive grouped choice problems in up to three stages. At the first stage of each round, subjects were graphically presented with a budget constraint $\mathbf{p}^1\mathbf{x} = 100$, where $\mathbf{p}^1 = (p_A^1, p_B^1)$ and $\mathbf{x} = (x_A, x_B)$. The price vector $\mathbf{p}^1$ was taken from the eight-point set

$$P = \{(1, 1.5), (2, 1), (1, 2.5), (3, 1), (1.5, 2), (2.5, 1.5), (3, 1.5), (2, 3)\}$$

of price vectors. Furthermore, the probability π of state A being chosen by a pseudo-random number generator was either 0.5 or 0.67.

All subjects were eventually presented with the complete set of all possible 16 first-stage choice problems which can result from combining one of the eight possible price vectors with one of the two probability distributions. The 16 possibilities were presented in random order, however.

2.3 Second Stage

Figure 4 shows how each subject's first-stage choice was used to construct the second-stage choice problem. The dashed line represents the first-stage budget line; the subject's portfolio choice is marked by one of the two dots — e.g., $\mathbf{x}^1 = (22, 67)$ in the figure. The subject, however, was shown only the second-stage budget line $\mathbf{p}^2\mathbf{x} = 100$. This was determined by first interchanging the two components of the first-stage price vector $\mathbf{p}^1$, then replacing the new higher component with a different one chosen at random. Specifically, if the first-stage price p_B^1 was lower, as in figure 4, and if x_B^1 denotes the amount allocated to asset B at the first-stage, then the second-stage price p_B^2 was determined by making a random choice from a uniform distribution on the closed interval $[100/x_B^1, 200/x_B^1]$, then rounding the result to the first decimal place. In the figure, we have $\mathbf{p}^2 = (1, 1.6)$.

In several cases, however, subjects chose first-stage portfolios that are stochastically dominated because, as discussed in Section 2.1, they lie to the left of the safe portfolio depicted in Figure 3. Worse still, in some cases subjects chose portfolios so close to the extreme where the whole budget is allocated to one security that our procedure would fail to determine a sensible second-stage choice problem, because the respective budget line would have had to be very steep (or flat). Our software, therefore, did not allow the subject to proceed beyond the first stage in case: either (i) the first-stage

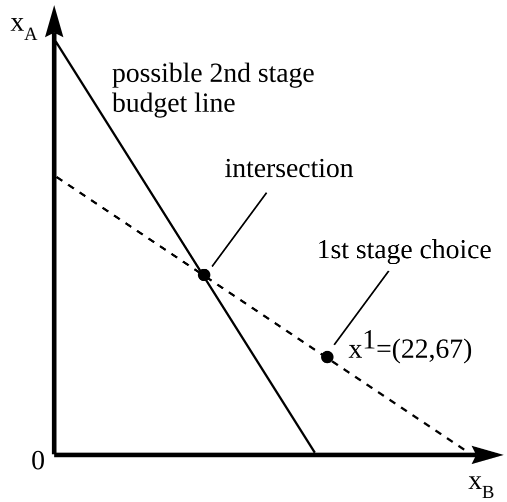

Figure 4: A second-stage choice problem with $p_A = 1$, $p_B = 1.6$, $\pi = 0.5$

choice was stochastically dominated; or (ii) the second-stage choice problem would have involved a price ratio greater than 10 (or smaller than 0.1).

2.4 Third Stage

Even if the subject had been allowed to proceed to the second stage, the portfolio chosen at this second stage could still fail to lie on the segment of the budget line between: (i) the extreme portfolio with $x_B = 0$; (ii) the intersection of the two budget lines as depicted in Figure 4. Indeed, if the subject's second-stage choice was either stochastically dominated or on the wrong side of the intersection marked in Figure 4, the computer program would omit the third stage and, unless all 16 rounds had already been completed, proceed directly to the next round in the sequence of three-stage experiments. Otherwise, as Figure 5 indicates, the third-stage budget

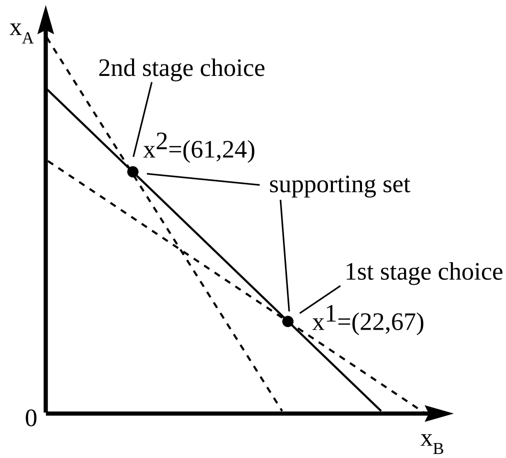

Figure 5: A third-stage choice problem with $p_A = 1.2$, $p_B = 1.1$, $\pi = 0.5$

constraint was constructed by taking the line through the different actual choices in the first two stages, then rounding both prices to the first decimal place. For example, assuming that the subject chose $\mathbf{x}^1 = (22, 67)$ at the first stage, followed by $\mathbf{x}^2 = (61, 24)$ at the second stage, the third-stage price vector would be $\mathbf{p}^3 = (1.2, 1.1)$ as indicated in Figure 5. Then Varian's supporting set consists of the line segment joining the first and second-stage portfolios.

2.5 Background

The experiment was conducted at the University of Warwick on 20th May, 2008, in a computer Laboratory that had often been used for experiments by other researchers. To avoid bias due to expert knowledge, we recruited 41 non-economics undergraduates (26 male and 15 female students had responded to our invitation before the deadline). All had previously agreed to be included a database of potential recruits for economic laboratory experiments and so were contacted by email.

The experiment was fully computerised. Standard software toolboxes in experimental economics and psychology such as z-Tree (Fischbacher, 2007) and Mouselab (Johnson et al., 1986) do not offer the graphical displays and the data structure we required. Instead, our experiment was programmed in Visual Basic.

Upon entering the laboratory, subjects were first given the on-screen instructions reproduced in the Appendix. Then a training session began where subjects were presented random budget lines and could make choices as often as they wanted. In order to end the training session and start the experiment, the subject had to click a button. This initiated a short countdown, after which the first-stage choice problem of the first round was displayed.

After each subject's last choice of the 16th round, the computer determined the amount they were owed, which was paid in cash. Everyone attending and completing the experiment was given £5 of UK currency. In addition, following the random lottery incentive scheme, subjects were told that one of the choice problems they were going to be presented would be randomly selected for an actual payment at the end of the experiment. The sum of all the payments was £461.20, which works out on average to £11.25 per participant, including the £5 participation fee.

3 A Nonparametric Statistical Test

3.1 Two Hypotheses: GARP and Random Choice

Bronars (1987) was concerned to show how GARP was a refutable hypothesis, even with the kind of aggregate data that Varian had considered. Accordingly, he had GARP as the null hypothesis, with Becker's (1962) model of uniformly random choice from the relevant budget line segment as a very specific alternative.

Instead, our concern will be to refute Becker's model of irrationality, where possible, by showing that it cannot explain the high proportion of observed choices satisfying GARP. Accordingly, our null hypothesis for each experimental subject is that, throughout the course of the experiment, a portfolio was always randomly selected from a uniform distribution over the current budget line segment; moreover, the random choices from successive budget lines are stochastically independent. Ignoring complications due to rounding, the probability of satisfying GARP in any one choice experiment is therefore the ratio of the length of the supporting set (the respective line segment in Figure 5) to the total length of the budget line segment.

3.2 Implications of Uniform Randomness

Formally, let the discretised budget set of the typical ith third-stage choice problem have K_i discrete elements ($i \in \{1, \ldots, I\}$), of which exactly k_i would satisfy GARP if chosen. Under the null hypothesis, the proportion $\kappa_i = k_i/K_i$ is the probability that the subject's randomly chosen portfolio satisfies GARP. Given a set Γ of I second or third-stage choice problems (up to a maximum of 16), there are 2^I ($\leq 2^{16} = 65,536$) different possible choice patterns of GARP compliance and noncompliance. Let H denote the set of all these 2^I possible patterns, and $G \subseteq H$ the subset of the I choice problems in which the subject's choices comply with GARP. Under the null hypothesis, each choice pattern $\gamma \in H$ occurs with probability

$$p_\gamma = \prod_{i \in G} \kappa_i \times \prod_{i \in I \setminus H} (1 - \kappa_i).$$

For each integer $\ell \in \{0, 1, \ldots, I\}$, let $H(\ell) \subset H$ denote the set of choice patterns that include exactly ℓ choices that are GARP consistent, and $I - \ell$ that are not. Then the probability of a subject exhibiting exactly ℓ GARP consistent choices is $P_\ell = \sum_{\gamma \in H(\ell)} p_\gamma$. Cumulating downwards gives, for each integer $z \in \{0, 1, \ldots, I\}$, the probability $1 - F(z) = \sum_{\ell=z}^{I} P_\ell$ that $\ell \geq z$.

3.3 Significance Tests

Let s denote the desired significance level of the test for GARP — for example, 5%. Let z_s denote the smallest possible integer satisfying $1 - F(z_s) \leq s$. Then we reject the null hypothesis of uniform randomness at the significance level s provided that the subject's choice pattern satisfies GARP on at least z_s occasions.

In principle the critical proportion $F(z_s)$ needed for this test could be calculated exactly from the finite stochastic process implied by the null hypothesis. In practice we used an obvious Monte Carlo simulation procedure to estimate $F(z_s)$ for each of the 11 particular values

$$s \in \{0.01, 0.05, 0.1, 0.2, 0.3, \ldots, 0.8, 0.9\}.$$

The dashed curve with squares as markers for the different significance levels in Figure 6 displays the results of 1000 simulations, which were enough for the observed proportions to converge. Of course, rounding implies the exact probability P_s that $F(\ell) \geq s$ will exceed s unless s is chosen exactly equal to one of the probabilities $F(z)$ for $z \in \{0, 1, \ldots, I\}$; this explains why the curve lies below the 45° line except at the end points $s = 0$ and $s = 1$. For this reason, our test slightly favours the null hypothesis of random choice.[1]

4 Experimental Results

4.1 Test Statistics

Table 1 gives an overview of our experimental results. Subjects were faced with an identical set of 16 first-stage choice problems, though the order in which each subject faced them was selected at random. No subject breached the time constraint of 30 seconds in any choice problem. Hence, in principle, there could have been 16 second-stage choices. But for reasons explained in Section 2.3, our procedure stopped after the first-stage choice if that was too inferior. On average, there were 4.2 such instances per subject, leaving us with a mean of 11.8 second-stage choices per subject — or 73.9% of the possible 16. Of the inferior choices, 84% (3.5 per subject) were dominated;

[1]There is an exact test with non-integer critical proportions $\hat{z}_s$ satisfying $F(\hat{z}_s) = s$ which takes the following form. Having found the integer critical value z_s as in the main text, classify as rational not only the subjects whose choice patterns satisfy GARP on at least z_s occasions, but also a random sample of those that satisfy GARP on $z_s - 1$ occasions where, independently of the others, each subject is included with probability $[s - F(z_s - 1)]/[F(z_s) - F(z_s - 1)]$.

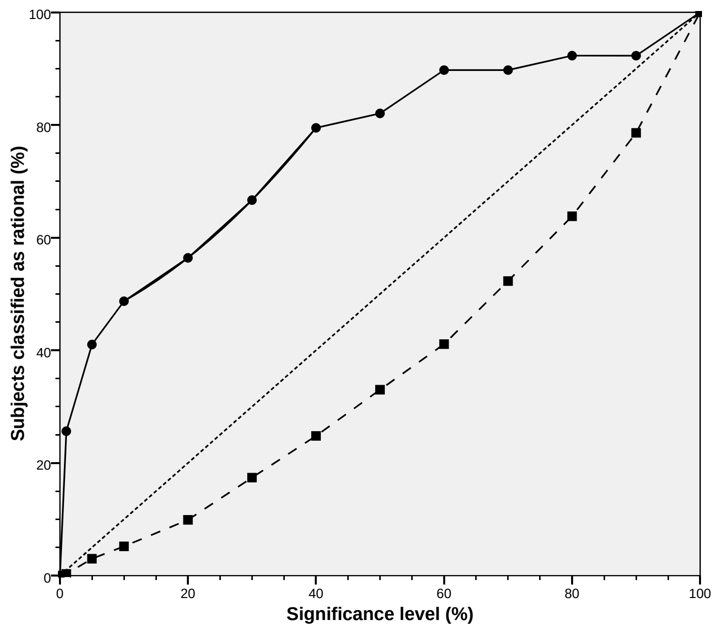

Figure 6: Test properties

Legend: The circular dots connected by solid lines represent actual observations of 39 subjects; the square dots connected by dashed lines represent a simulation using 1000 random players; the dotted line represents the theoretical level of significance.

the remaining 16% (0.7 per subject) were extreme in the sense that, if we had continued with our algorithm, it would have required the price ratio p_A^2/p_B^2 to be either larger than 10 or smaller than 0.1. About 58.6% of the second-stage choices (6.9 per subject) enabled us to construct a third-stage choice. The remaining 4.9 second-stage choices per subject were either dominated (1.9 per subject) or (3.0 per subject) on the wrong side of the point where the first-stage and the second-stage budget lines intersect, thus making it impossible to construct the relevant supporting set. Finally, 5.4 third-stage choices per subject (77.5% of all third-stage choices) were both undominated and consistent with GARP. The remaining 1.6 third-stage choices either violated GARP (1.4 per subject) or were stochastically dominated (0.2).

Table 1: GARP Consistency of Choices: Aggregated Data

Number of consistent choices	Stage 1	Stage 2	Stage 3
theoretical maximum	16	16	16
mean number of consistent choices	11.8	6.9	5.4
mean % of maximum	73.9%	43.1%	33.8%
mean % of previous column	–	58.6%	77.5%

$N = 41$ *subjects.*

Any test of individual rationality requires disaggregated data. After all, the fact that about 78% of all third-stage choices were GARP consistent says little about how consistent each individual's choices were. Table 2 lists each subject's ID in the experiment (for reference purposes only), followed by statistics concerning their performance in the third-stage choice problems. For the subjects with ID numbers in the range 1–21, columns 2–5 respectively report the total number of third-stage choices I, then the number z and proportion z/I of GARP consistent and undominated third-stage choices, followed by the significance level $p(z)$ of our rationality test. Columns 7–10 do the same for the subjects with ID numbers in the range 22–41.

The p-values that are reported in columns 5 and 11 of Table 2 are computed to allow a separate non-parametric exact test for each subject, based on all possible permutations of choice patterns. They specify the conditional probability that a third-stage choice satisfies GARP, given that the subject's first and second stage choices did not rule out reaching the third stage.

4.2 Discussion of Results

Subject 13, for example, was a male who got to the third stage in 13 out of the 16 maximum possible grouped choice problems. Of these 13, no less than 12 (92%) third-stage choices lay within the support sets for GARP and were undominated. The probability that his random choices would pass the test at least 12 out of 13 times is just 0.007%. Hence, using our significance level of 10%, we reject the null hypothesis that his 12 GARP consistent choices were purely random. Note that two female subjects (#7 and #12) did not reach any third-stage choice essentially because they always chose only stochastically dominated portfolios in the first or second stage. The following discussion focuses only on the remaining 39 subjects who reached the third stage at least once.

Figure 6 displays the results graphically. Although our subjects often violated GARP, they did so distinctly less often than the uniformly random consumer would have done. For example, at the 10% significance level, 19 subjects (or 48.7%) were classified as rational. Comparing the curves for stage three and for the simulation shows that our test is powerful enough to distinguish clearly between: (i) actual subjects who could validly be classified as rational; (ii) simulated random subjects that were incorrectly classified as rational. It is worth recalling that Sippel's (1997) procedure with an Afriat choice efficiency index of 0.9 classified as rational no fewer than 98.5% of the

Table 2: GARP Consistency of Choices: Individual Data

ID	total	consistent		p-value†	ID	total	consistent		p-value†
	I	z	z/I	$p(z)$		I	z	z/I	$p(z)$
1	6	5	0.83	0.050*	22	7	4	0.57	0.079*
2	11	7	0.64	0.219	23	6	3	0.50	0.316
3	3	3	1.00	0.032*	24	10	8	0.80	0.114
4	8	7	0.88	0.003*	25	11	8	0.73	0.244
5	5	4	0.80	0.128	26	14	12	0.86	0.062*
6	15	15	1.00	0.000*	27	3	2	0.67	0.355
7	0	0	—	—	28	5	3	0.60	0.069*
8	11	10	0.91	0.003*	29	4	3	0.75	0.173
9	13	12	0.92	0.033*	30	2	0	0.00	1.000
10	10	9	0.90	0.002*	31	10	9	0.90	0.000*
11	0	0	—	—	32	10	9	0.90	0.001*
12	3	2	0.67	0.381	33	6	3	0.50	0.546
13	13	12	0.92	0.007*	34	12	10	0.83	0.008*
14	3	3	1.00	0.045*	35	3	2	0.67	0.405
15	1	1	1.00	0.370	36	7	5	0.71	0.042*
16	14	9	0.64	0.249	37	11	10	0.91	0.003*
17	4	2	0.50	0.261	38	7	5	0.71	0.042*
18	3	2	0.67	0.333	39	6	6	1.00	0.003*
19	1	0	0.00	1.000	40	10	4	0.40	0.782
20	1	0	0.00	1.000	41	2	1	0.50	0.565
21	13	10	0.77	0.512					

†*Significance level of a non-parametric exact test. The null hypothesis is random choice. An asterisk indicates significance at the 10% level.*

randomly generated consumers (as against 91.7% of the actual subjects).[2] By contrast, our three-stage test classified only 5.2% of simulated subjects as rational.

Finally, Table 3 reports the results of some tests for gender differences. The share of GARP consistent choices was significantly greater for male subjects. Likewise, the mean rejection probability reported in Table 2 was much higher for female subjects. A possible explanation for this may lie in the relative shares of portfolios chosen by each gender that were first-order stochastically dominated, as reported in the last three rows of Table 3. In all three stages female subjects chose between two and four times as many dominated portfolios as their male counterparts. This may reflect a higher proportion of male subjects with some prior experience of the type of investment problem and of graphical computer display that was used in the experiment.

Table 4 gives the results of a pooled-sample logit regression whose dependent indicator variable equals 1 if and only if the subject chose a first-order stochastically dominated portfolio. To allow for the panel structure of the data, the regression used a method of robust covariance estimation. The exogenous variables are gender, round, and the interaction term between these two. We conducted the regression first for the whole sample, then for each stage separately, both with and without the interaction term. The table shows that the gender term is highly significant for all regressions except the

[2]Following Bronars' (1987) proposal, Sippel actually compared his experimental data with a set of 1000 demand vectors created using randomly determined constant budget shares, which correspond to Cobb–Douglas preferences — for details, see Sippel (1996).

Table 3: GARP Consistency of Individual Choices: Gender Differences

	Gender				Significance
	female		male		level
	mean	s.e.	mean	s.e.	
share of GARP consistent choices					
3rd stage	0.647	(0.058)	0.815	(0.026)	0.010*
probability of rejecting substantive rationality					
3rd stage	0.381	(0.083)	0.164	(0.052)	0.026*
share of dominated portfolios					
1st stage	0.324	(0.029)	0.153	(0.018)	0.000*
2nd stage	0.256	(0.035)	0.108	(0.017)	0.000*
3rd stage	0.162	(0.045)	0.042	(0.014)	0.013*

**Significant at the 10%-level according to a two-tailed independent-sample t test (checked for equality of variances).*

one for the third stage when an interaction term is included. This accords with our previous result that, in general, female subjects chose dominated portfolios more often than males. Also, there were no significant round or learning effects, nor any significant interaction between round and gender.

5 Conclusion

We have reported and analysed an experiment in which subjects were faced with a series of 16 grouped three-stage portfolio-selection problems. In previous studies significance levels were computed by tolerating small changes of the chosen portfolios or budget lines — that is, they allow an Afriat efficiency index below one. In contrast, for our test to be passed, a sufficient number of portfolios chosen in a sequence of three-stage problems have to satisfy the relevant inequalities *exactly*. This sharpens the distinction between truly rational subjects and random players.

Overall, using a 10% significance level, only 19 out of 39 subjects (or 48.7%) could be classified as having made third-stage choices that all passed our test of the standard GARP axiom of revealed-preference theory. Even though our nonparametric test is easier to satisfy than many predecessors, this proportion is distinctly lower than in previous studies. This may be due in part to the fact that economics undergraduates subjects were specifically excluded from our subject pool.

Table 4: Gender Differences — Logit Estimates

	Coefficient							
Variable	All Choices		1st Stage		2nd Stage		3rd Stage	
Intercept	-2.045***	-2.266***	-1.663***	-1.805***	-2.041***	-2.412***	-3.428***	-3.536***
	(0.167)	(0.223)	(0.224)	(0.296)	(0.290)	(0.393)	(0.578)	(0.770)
Gender	1.125***	1.555***	0.981***	1.251***	1.050***	1.787***	1.478**	1.687
($Female = 1$)	(0.145)	(0.306)	(0.193)	(0.405)	(0.255)	(0.537)	(0.474)	(1.064)
Round	-0.004	0.022	-0.006	0.010	-0.009	0.034	0.034	0.045
	(0.016)	(0.022)	(0.021)	(0.030)	(0.027)	(0.039)	(0.051)	(0.074)
Gender	—	-0.050	—	-0.032	—	-0.087	—	-0.027
× Round		(0.031)		(0.042)		(0.055)		(0.103)
n	1425	1425	656	656	485	485	284	284
LL	-615.645	-614.355	-332.050	-331.757	-202.045	-200.792	-67.294	-67.270
Pseudo-R^2	0.047	0.049	0.038	0.039	0.040	0.046	0.070	0.070

Table notes. Dependent variable: dominated portfolio chosen. Binary logit with robust covariance matrix estimation. First line: coefficients; second line: standard errors. *$p \leq 0.10$, **$p \leq 0.05$, ***$p \leq 0.01$.

Acknowledgments

Generous support for Peter Hammond from a Marie Curie Excellence Chair funded by the European Commission under contract number MEXC-CT-2006-041121 is gratefully acknowledged, as is assistance from the Institute of Advanced Study at the University of Warwick that facilitated a visit by Stefan Traub during which the experiment was performed. Both authors would also like to thank Malena Digiuni for indispensable assistance in arranging the experiment, as well as seminar audiences at the University of Warwick, and also Thomas Palfrey, for encouraging comments and suggestions.

References

Afriat, S.N. (1973): "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method", *International Economic Review* 14, 460–472.

Andreoni, J., and Harbaugh, W.T. (2008): *Power Indices for Revealed Preference Tests*, University of Wisconsin Madison Working Paper; now available at `http://econ.ucsd.edu/~jandreon/WorkingPapers/power.pdf`.

Becker, G.S. (1962): "Irrational Behavior and Economic Theory", *Journal of Political Economy* 70, 1–13.

Bronars, S.G. (1987): "The Power of Nonparametric Tests of Preference Maximization", *Econometrica* 66, 693–698.

Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007a): "Revealing Preferences Graphically: An Old Method Gets a New Tool Kit", *American Economic Review* 97, 153–158.

Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007b): "Consistency and Heterogeneity of Individual Behavior under Uncertainty", *American Economic Review* 97, 1858–1876.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Ready-made Economic Experiments", *Experimental Economics* 10, 171–178.

Johnson, Eric J., Payne, John W., Bettman, James R., and Schkade, David A. (1986): *Johnson Monitoring Information Processing and Decisions: The Mouselab System*, Technical Report, Duke University.

Samuelson, P.A. (1938): "A Note on the Pure Theory of Consumer's Behaviour", *Economica* 5, 61–71.

Sippel, R. (1996): "A Note on the Power of Revealed Preferences Tests with Afriat Inefficiency", SFB 303 Discussion Paper A-528, Universität Bonn, Bonn.

Sippel, R. (1997): "An Experiment on the Pure Theory of Consumer's Behaviour", *Economic Journal* 107, 1431–1444.

Varian, H.R. (1982): "The Nonparametric Approach to Demand Analysis", *Econometrica* 50, 945–973.

Varian, H.R. (2006): "Revealed preference", in: M. Szenberg, L. Ramrattan, A.A. Gottesman (eds.): *Samuelsonian Economics and the Twenty-first Century*, Oxford University Press, Oxford, 99–115.

Appendix

Instructions

Experimental Instructions (Please, read carefully)

This is an experiment in decision-making. The entire experiment should be complete within about 30 minutes. Research foundations have provided funds for conducting this research. Please, pay careful attention to the instructions as a considerable amount of money is at stake. At the end of the experiment, you will be paid privately. Your payoffs will depend partly on your decisions and partly on chance, but not on the decisions of the other participants in the experiment. You will receive 5 pounds as a participation fee. In addition you will receive a payment whose calculation will be explained in the following.

During the experiment, we will speak in terms of experimental "tokens" instead of pounds. At the end of the experiment your payoff will be calculated in tokens and translated into pounds. The exchange rate between tokens and pounds is stated on a note at your workplace.

In each decision problem, you will be asked to allocate an initial endowment of 100 tokens between two accounts labeled A and B. The A account corresponds to the vertical and the B account to the horizontal axis in a two-dimensional graph. Each choice will involve choosing with the mouse pointer a point on a blue line representing possible token allocations. In each choice, you may choose any A and B pair that is on the blue line.

Each decision problem will start by having the computer select such a line randomly, where each line permits a minimum of 10 and a maximum of 100 tokens on each account. The "prices" for the two accounts are stated on *the right side of the screen. An example: the blue line runs from 50 on the vertical axis (account A) to 33 on the horizontal axis (account B). Hence, the price for allocating a token to account A is two tokens, and for a token on account B you have to give up three tokens of your initial endowment.*

You have exactly 30 seconds for choosing one point on the blue line. The time remaining is stated on the screen. Furthermore, you will receive an acoustic signal during the last five seconds.

To choose an allocation, use the mouse to move the pointer over the blue line. You will be shown the token allocations that belong to the respective points on the blue line. Once you have found the allocation that you like best, click with the left mouse button somewhere on the screen, and the most recent allocation will be fixed. If you want to revise your decision, click the left mouse button again and the line will be released. If you are satisfied with your decision, click the "OK" button with the mouse pointer.

As noted above, you can choose only allocations that are located on the blue line. You have 30 seconds for each choice. If you run out of time before you fixed an allocation, the computer will automatically move on to the next decision problem. If you did not touch the blue line at least once within the 30 seconds in order to display an allocation, the computer will record that you did not make a decision; if you displayed an allocation but did not fix it by mouse click, the computer will record the most recent allocation as your choice. You cannot revise your decision after having clicked the "OK" button or the 30 seconds have elapsed.

Afterwards you are asked for your next decision. At the end you will be informed that the experiment has ended and the computer determines you payoff.

Your payoff is determined as follows: at the end of the experiment the computer will randomly select one decision round. It is equally likely that any round will be chosen. Afterwards the computer will decide whether account A or B will be paid off. The probability of an account to be selected is stated on the screen for each decision problem. The probability is either 50:50 or 67:33. Pay attention to the probabilities shown on the screen while making your choice. At the beginning of each decision problem, the probabilities briefly flash up in red color. Be careful: if the computer selects a decision task in which you did not make a choice, your payoff will be zero.

Your payoff in tokens, your choice, and the account that has been selected, will be shown in a popup window. Please, let our assistant know that you have finished.

Your participation in the experiment, your choices, and your payoff will be kept confidential. Only on the payoff receipt will we have to record your name. In order to keep your privacy you should not talk to anyone about the experiment and your choices (at least until the complete experiment has ended). We would like to ask you not to talk during the experiment and to remain silent until the end of the last round.

If you are ready for a trial run, click the "OK" button. If there are open questions, please, contact one of our assistants.