A Three-Stage Experimental Test of Revealed Preference

Peter Hammond, Stefan Traub

Department of Economics, University of Warwick, Coventry, UK
Department of Economics & ZeS, University of Bremen, Germany

9th November, 2010

Abstract A new direct three-stage experimental test of Varian’s (1982) generalised axiom of revealed preference (GARP) is applied to 41 different subjects’ choices in a series of 16 successive grouped portfolio selection problems. Almost 80% of our subjects are observed to violate GARP, even by our more generous nonparametric test. However, most subjects who satisfied GARP at a second stage of the test also satisfied it at the succeeding third stage. In general, the proportion of choices satisfying GARP was significantly higher for male rather than female subjects.

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 pardon even the slightest inconsistency; any corresponding 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) 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 $x^1$ when the price vector was $p^1$. By definition $x^1$ is revealed preferred to any alternative bundle $x^2$ satisfying $p^1 x^2 < p^1 x^1$. Suppose nevertheless that the same consumer were also observed choosing the bundle $x^2$ when the price vector is $p^2$, where $p^2 x^2 > p^2 x^1$. This would imply that $x^2$ is revealed preferred to $x^1$, and so violate GARP. The Afriat efficiency index of the choice $x^2$ is the ratio $p^1 x^2 / p^1 x^1$, which is evidently less than 1.

Allowing choices whose efficiency index is less than one relaxes the GARP axiom, and so leads to a considerable increase in subjects’ measured compliance 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 $p^1 = (1.25, 1)$ and $p^2 = (1, 1.25)$, as illustrated in Fig. 1. Assume too that at prices $p^1$ a person chooses the consumption bundle $x^1 = (x_A^1, x_B^1) = (64, 20)$, or indeed any other bundle on the line segment joining the end point $Q$ to the point $P = (44\frac{4}{5}, 44\frac{4}{5}) \approx (44.4, 44.4)$ where the two budget lines intersect. Then it is straightforward to show that at prices $p^2$ 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. Allowing an Afriat efficiency index of 0.9, however, which is equivalent to throwing away $10 at prices p^2, moves the intersection of
Figure 1: Basic Example
the two budget lines down to the point \( P' = (22\frac{2}{3}, 62\frac{2}{3}) \approx (22.2, 62.2) \). This extends the supporting set to the line segment \( P'Q' \), so the chance of a random choice being classified as rational rises to \( \frac{36}{51} \approx 69.1\% \).

Recently Andreoni and Harbaugh (2008) have extensively discussed the pros and cons of several different power indices for revealed preference tests, including that of Bronars (1987). We refer the interested reader to their paper for more details, and proceed directly to presenting our own novel experimental design allowing a more direct test.

### 1.2 A Three-Stage Direct Test

Consider any list \( s^n = (p^i, x^i)_{i=1}^n \) of \( n \) pairs of price and quantity vectors that satisfy both GARP and the normalization \( p^i x^i = 1 \) \((i = 1, \ldots, n)\). Let \( p^{n+1} \) be any previously unobserved price vector. Then Varian (1982, 2006) defines the supporting set \( S(p^{n+1}; s^n) \) of consumption bundles \( x^{n+1} \) as those for which the extended sequence \( (p^i, x^i)_{i=1}^{n+1} \) also satisfies both GARP and the normalization \( p^i x^i = 1 \) \((i = 1, \ldots, n + 1)\). 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 \( x^1 \) 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 consumer’s new demand $\mathbf{x}^2$ 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, there was 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 the 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 $\langle x_A, x_B \rangle$ of nonnegative integers immediately below the budget line. Specifically, we allowed any nonnegative integer allocation satisfying

$$100 - \max\{p_A, p_B\} < p_A x_A + p_B x_B \leq 100.$$
Figure 2: Example screen
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 moved close enough to a feasible allocation, that allocation was indicated by two numbers and by associated reference lines marked in red. As long as the mouse pointer was not moved too far away from this position, this information remained visible, but it disappeared from the screen as soon as the mouse pointer was moved far enough away. 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: A first-stage choice problem with $p_A = 1.5$, $p_B = 1$, $\pi = 0.5$
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 $45^\circ$-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.

### 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 $p^1 x = 100$, where $p^1 = (p^1_A, p^1_B)$ and $x = (x_A, x_B)$. The price vector $p^1$ 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 $\pi$ 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. They 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 and the dot marks the subject’s portfolio choice — e.g., \( x^1 = (22, 67) \) in the figure. Neither of these was shown to the subject, who only saw the second-stage budget line \( p^2 x = 100 \). This was determined by first interchanging the two components of the first-stage price vector \( p^1 \), then replacing the new higher component with a different one chosen at random. Specifically, if the first-stage price \( p^1_B \) was lower, as in figure 4, and if \( x^1_B \) denotes the first-stage allocation to asset \( B \), then the second-stage price \( p^2_B \) was determined by making a random choice from a uniform distribution on the closed interval \([100/x^1_B, 200/x^1_B]\), then rounding the result to the first decimal place. In the figure, we have \( p^2 = (1, 1.6) \).

In several cases, however, subjects chose dominated portfolios close to the extreme where the whole budget is allocated to the more costly security. In these cases our procedure ruled out a sensible second-stage choice problem, because the respective budget line would have had to be very steep (or flat). Our software, therefore, was programmed so that the subject did not proceed beyond the first stage in case the second-stage choice problem would have involved a price ratio greater than 10 (or smaller than 0.1).

After the indicated first stage choice, Varian’s supporting set for the subject’s second stage-choice consists of the segment of the budget line between the extreme portfolio with \( x_B = 0 \) and the intersection of the two budget lines.
Figure 4: A second-stage choice problem with $p_A = 1$, $p_B = 1.6$, $\pi = 0.5$
2.4 Third Stage

If a subject’s second-stage choice was from the supporting set, the subject was faced with a third-stage problem. Otherwise, the computer program omitted the third stage and, unless the 16 rounds had already been completed, proceeded directly to the next round in the sequence of three-stage experiments.

As Figure 5 indicates, the third-stage budget constraint was constructed by taking the line that would pass through the two different allocations that were actually chosen in the first two stages, then rounding both prices to the first decimal place. For example, assuming that the subject chose $x^1 = (22, 67)$ at the first stage, followed by $x^2 = (61, 24)$ at the second stage, the third-stage price vector would be $p^3 = (1.2, 1.1)$ 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 in time). All had previously agreed to be included a database of potential recruits for economic laboratory experiments and so were contacted by email. Everyone attending and completing the experiment was given £5 of UK currency. In addition, following the random lottery in-
Figure 5: A third-stage choice problem with $p_A = 1.2$, $p_B = 1.1$, $\pi = 0.5$
centive 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 experiment was fully computerised. Since 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 required for our experiment, it 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 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. 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, we take uniform randomness as our null hypothesis.

More precisely, we consider a representative irrational subject who throughout the course of the experiment always chooses a portfolio randomly 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 line segment in Figure 4 or 5, as appropriate, to the total length of the budget line segment.

### 3.2 Implications of Uniform Randomness

Formally, let the discretised budget set of the typical $i$th second or 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 $\Gamma$ 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 show exactly $\ell$ choices that are GARP consistent, and $I - \ell$ that are not. Then the probability of a subject exhibiting exactly $\ell$ 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

Suppose the desired significance level of the test for GARP is $s$ — 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 solid curve in figure 6 displays the results of 1000 simulations, which was 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 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.\(^1\)

---

\(^1\)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
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 1.4 such instances per subject, leaving us with a mean of 14.6 second-stage choices per subject — or 91% of the possible 16. About 55% of all second-stage choices (8.1 per subject) were GARP consistent.

When the second-stage choice was GARP consistent, a third-stage choice problem was constructed according to the procedure explained in Section 2.4. Of these third-stage choices, Table 1 reports that 78% were GARP consistent — that is, correctly predicted by the supporting set.

Any test of individual rationality requires disaggregated data. After all, the fact that about 55% of all second-stage choices were GARP consistent does not say very much about how consistent any individual might have been. Table 2 lists each subject’s ID in the experiment (for reference purposes only), followed by statistics concerning their performance in the second and third-stage choice problems. Columns 2–5 specify respectively the total number of second-stage choices $I_2$, then the number $z_2$ and proportion $z_2/I_2$ of GARP consistent second-stage choices, followed by the significance level $p(z_2)$ of our
Table 1: GARP Consistency of Choices: Aggregated Data

<table>
<thead>
<tr>
<th></th>
<th>Outset</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st</td>
<td>2nd</td>
</tr>
<tr>
<td>maximum number of choices</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>actual number of choices</td>
<td>16</td>
<td>14.6</td>
</tr>
<tr>
<td>actual number (%)</td>
<td>100</td>
<td>91</td>
</tr>
<tr>
<td>% of previous column</td>
<td>—</td>
<td>91</td>
</tr>
</tbody>
</table>

rationality test. The last three columns state the same data for the third-stage choices, except that $I_3$ can be omitted because it is equal to $z_2$.

Columns 5 and 8 of Table 2 report $p$-values that are computed to allow a separate non-parametric exact test for each subject, based on all possible permutations of choice patterns. The second and third-stage tests, however, have slightly different interpretations. The figure in column 5 concerns the conditional probability of a second-stage choice satisfying GARP, given that the subjects’ first-stage choice did not rule out reaching the second stage. Similarly, the figure in column 8 concerns the conditional probability of a third-stage choice satisfying GARP, given that the subjects’ choices at the first two stages allowed the third stage to be reached. In particular, the latter test gives the conditional probability of the third-round choice satisfying GARP given that the previous second-stage choice was already GARP consistent.
Table 2: GARP Consistency of Choices: Individual Data

| Subject | 2nd-stage Choices | | | 3rd-stage Choices | | |
|---------|-------------------|------------------|------------------|
|         | total | GARP consistent | p-value* | GARP consistent | p-value* |
|         | $I_2$ | $z_2$ | $z_2/I_2$ | $p(z_2)$ | $z_3$ | $z_3/I_3$ | $p(z_3)$ |
| 1 | 16 | 8 | 0.50 | 0.829 | 7 | 0.88 | 0.007* |
| 2 | 16 | 15 | 0.94 | 0.001* | 11 | 0.73 | 0.030* |
| 3 | 14 | 3 | 0.21 | 0.997 | 3 | 1.00 | 0.031* |
| 4 | 16 | 9 | 0.56 | 0.538 | 7 | 0.78 | 0.016* |
| 5 | 15 | 7 | 0.47 | 0.975 | 6 | 0.86 | 0.043* |
| 6 | 16 | 16 | 1.00 | 0.000* | 15 | 0.94 | 0.000* |
| 7 | 15 | 0 | 0.00 | 1.000 | — | — | — |
| 8 | 13 | 13 | 1.00 | 0.052* | 12 | 0.92 | 0.000* |
| 9 | 15 | 15 | 1.00 | 0.004* | 13 | 0.87 | 0.054* |
| 10 | 16 | 10 | 0.63 | 0.419 | 9 | 0.90 | 0.002* |
| 11 | 16 | 1 | 0.06 | 1.000 | 0 | 0.00 | 1.000 |
| 12 | 12 | 8 | 0.67 | 0.284 | 5 | 0.43 | 0.085* |
| 13 | 15 | 13 | 0.87 | 0.024* | 12 | 0.92 | 0.007* |
| 14 | 12 | 5 | 0.42 | 0.801 | 4 | 0.80 | 0.067* |
| 15 | 16 | 2 | 0.13 | 1.000 | 1 | 0.50 | 0.130 |
| 16 | 16 | 15 | 0.94 | 0.002* | 11 | 0.73 | 0.079* |
| 17 | 15 | 5 | 0.33 | 0.975 | 3 | 0.60 | 0.102 |
| 18 | 13 | 4 | 0.31 | 0.980 | 2 | 0.50 | 0.454 |
| 19 | 10 | 1 | 0.10 | 0.999 | 0 | 0.00 | 1.000 |
| 20 | 14 | 2 | 0.14 | 1.000 | 0 | 0.00 | 1.000 |
| 21 | 16 | 16 | 1.00 | 0.003* | 13 | 0.81 | 0.267 |

Table continues.
Continuation of Table 2

<table>
<thead>
<tr>
<th>Subject</th>
<th>2nd-stage Choices</th>
<th>3rd-stage Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ID</td>
<td>total</td>
</tr>
<tr>
<td>---------</td>
<td>----</td>
<td>-------</td>
</tr>
<tr>
<td>22</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>23</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>24</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>25</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>26</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>27</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>28</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>29</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>30</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>32</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>33</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>34</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>35</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>36</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>37</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>38</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>39</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>40</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>41</td>
<td>14</td>
<td>2</td>
</tr>
</tbody>
</table>

*Significance level of a non-parametric exact test based on all possible permutations of choice patterns. The null hypothesis is random prediction.
4.2 Discussion of Results

First, we comment on the second-stage choices. Subject 13, for example, was a male who was presented with 15 (out of 16 maximum possible) second-stage choice problems. Of these 13 (87%) lay within the support sets for GARP. The probability that his random choices would pass the test at least 13 out of 15 times is just 2.4%. Hence, using our significance level of 10%, we reject the null hypothesis that his 13 GARP consistent choices were purely random.

Overall, using a 10% significance level, only 8 out of 41 subjects (or 19.5%) passed our test for all their second and third stage choices. Even some of these made one or more stochastically dominated choices at the first stage. This proportion is distinctly lower than in previous studies where, however, significance levels were computed by tolerating small changes in the chosen portfolios or budget lines. In contrast, our three-stage test requires an appropriate proportion of observed demands to satisfy GARP exactly. Nevertheless, the degree of rationality our subjects achieve seems remarkably low. This may be due in part to the fact that our subjects were non-economics undergraduates.

The GARP hypothesis fares rather better when restricted to third-stage choices. After omitting one subject who did not reach any third-stage task, there were 25 of the remaining 40 subjects — that is 62.5% — whose choices passed our test for third-stage rationality, conditional on having made rational second-stage choices. With one exception, all subjects who performed well in the second stage were successful in the third stage too.
that GARP consistency may be to a great extent decision-task as well as subject specific.

Figure 6 displays the results graphically. Our subjects’ second-stage performance was not too impressive, though distinctly better than for the uniformly random consumer. For example, at the 10% significance level, only 9 subjects (or 22%) were classified as rational. Despite this low proportion, comparing the curves for stage two 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 Sippe1’s (1997) procedure with an Afriat choice efficiency index of 0.9 classified as rational no fewer than 98.5% of the randomly generated consumers (as against 91.7% of the actual subjects).\footnote{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).} Compared to this, our three-stage test classified only 6.7% of simulated subjects as rational.

The same figure shows that our subjects’ third-stage choices were much more likely to satisfy GARP. No less than 61% of subjects passed our rationality test at the 10% significance level. Also, once again, we could clearly distinguish between actual and simulated subjects.

Finally, Table 3 reports the results of some tests for gender differences. The share of GARP consistent choices was significantly greater for male than for female subjects at both the second and third stages. Likewise the mean rejection probability reported in Table 2 was much higher for female
Table 3: GARP Consistency of Individual Choices: Gender Differences

<table>
<thead>
<tr>
<th>Stage</th>
<th>Gender</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>female (mean)</td>
<td>male (mean)</td>
</tr>
<tr>
<td></td>
<td>s.e.</td>
<td>s.e.</td>
</tr>
<tr>
<td>share of GARP consistent choices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd stage</td>
<td>0.359 (0.071)</td>
<td>0.668 (0.052)</td>
</tr>
<tr>
<td>3rd stage</td>
<td>0.474 (0.078)</td>
<td>0.764 (0.046)</td>
</tr>
<tr>
<td>rejection probability of substantive rationality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd stage</td>
<td>0.820 (0.083)</td>
<td>0.399 (0.075)</td>
</tr>
<tr>
<td>3rd stage</td>
<td>0.448 (0.089)</td>
<td>0.110 (0.047)</td>
</tr>
<tr>
<td>share of dominated portfolios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st stage</td>
<td>0.324 (0.049)</td>
<td>0.153 (0.032)</td>
</tr>
<tr>
<td>2nd stage</td>
<td>0.354 (0.056)</td>
<td>0.138 (0.039)</td>
</tr>
<tr>
<td>3rd stage</td>
<td>0.267 (0.087)</td>
<td>0.078 (0.029)</td>
</tr>
</tbody>
</table>

*Significant at the 10%-level. Two-tailed independent-sample t test (checked for equality of variances). n_{female} = 16, n_{male} = 25, except for the 3rd stage share of dominated portfolios, where n_{female} = 15.
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 the likelihood that male subjects brought along more outside-university experience with investment problems and with the type of graphical computer display we used in our experiment.

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. Only 22% of our subjects could be classified as making choices that satisfied the standard GARP axiom of revealed-preference theory, even though our non-parametric test is easier to satisfy than many predecessors. Conditional on making a rational choice at the second stage, however, 62.5% of our subjects went on to make a rational choice at the third stage as well. This suggests that rationality is to a large extent both task and subject specific.

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, our test considers a sequence of three-stage choice problems, of which a certain number has to be fulfilled exactly. This sharpens the distinction between truly rational subjects and random players.
Acknowledgments

Generous support from a Marie Curie Excellence Chair funded by the European Commission under contract number MEXC-CT-2006-041121 is gratefully acknowledged. The 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


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.