FULL-SCALE REAL TESTS OF CONSUMER BEHAVIOR
USING EXPERIMENTAL DATA

by

Aurelio MATTEI*

University of Lausanne

ABSTRACT

This paper reports the results of three experiments on the neoclassical theory of consumer behavior. All these experiments were performed with the aim to simulate as closely as possible the behavior of consumers in the real world. Real goods, real money and the same incentives as in the real world were used. The large number of participants (about 450 individuals) is another important characteristic of these experiments.

Nonparametric tests using the theory of revealed preference show that a significant number of individuals have an inconsistent behavior. If a nearly optimizing behavior is postulated then most of these inconsistencies disappear but a model of random choices gives almost the same results. The hypothesis of a representative consumer (average behavior) seems a more fruitful approach.

These experiments were designed so that both nonparametric and parametric tests could be used. A flexible functional form is estimated and the classical tests of homogeneity and symmetry show that there are no fundamental differences between the two types of test if some approximations are taken into account.

Key words: consumer theory, experimental economics, revealed preference, flexible functional forms, nonparametric and parametric tests

JEL classification: C91, D12

* Address: Prof. A. Mattei, DEEP, University of Lausanne, Ecole des HEC, CH-1015 Lausanne / Switzerland (e-mail:AMATTEI@HEC.UNIL.CH). Financial support from the Swiss National Foundation for Scientific Research is gratefully acknowledged. Laurent Cretegny and David Escher provided excellent research assistance.
I. Introduction

Controlled experiments of individual decision making under certainty are rare. This is very surprising given the central place of the neoclassical model of consumer behavior in economics. The absence of paradoxes is probably the main reason of the great difference with respect to the numerous experiments of decision making under uncertainty.

The conclusions of the first empirical tests of consumer behavior were generally favourable to the theory. However, later estimations with flexible functional forms rejected the homogeneity and symmetry conditions. Since the same results were obtained using different functional forms, the validity of the static neoclassical model of consumer behavior was questioned\(^1\) if not rejected\(^2\).

A simple test of consumer behavior can be performed using the theory of revealed preferences. Varian (1982) gives an efficient algorithm that can be applied to any amount of data. When this nonparametric test was applied to the data used for the estimation of flexible functional forms, the conclusions were very different. All these data\(^3\) could have been generated by a utility-maximizing consumer. Hence, the negative results with the flexible forms were exclusively due to the approximation used in obtaining these forms.

The conclusions of a test based on the theory of revealed preferences can be misleading if the budget sets do not intersect\(^4\). Any model which satisfies the budget constraint is consistent with the data. In particular, the data could have been generated by a random choice.

---

\(^1\) "... at this stage it is not necessary to abandon the axioms of choice in the face of the results of this chapter. Ultimately, of course, given sufficiently convincing evidence, we should be prepared to do so." Deaton and Muellbauer (1980, p. 82)

\(^2\) "Our results ... make possible an unambiguous rejection of the theory of demand." Christensen, Jorgenson, and Lau (1975, p. 381)

\(^3\) Per capita consumption expenditures taken from the national income accounts.

\(^4\) Varian was fully aware of this problem. See his remarks on page 965.
Becker (1962) assumes that an impulsive consumer chooses randomly a point on the budget hyperplan. In this case, all the points have "an equal chance of being selected". Bronars (1987) suggests to measure the power of the nonparametric test by generating random consumption data which exhaust the budget set. If all these data are consistent with preference maximization then the power of the test is nil.

Panel data are more suitable than time series of aggregate data to test the model of consumer behavior. Since the changes in relative prices are small, we need however a long period of time to obtain a test with a strong power\(^5\). But in this case one cannot exclude that tastes have changed and therefore it is impossible to discriminate between an inconsistent behavior and a change in tastes.

With controlled experiments, we can have a high power of the test and at the same time satisfy all the theoretical hypotheses. Artificial conditions shall be kept at a minimum in order to simulate as closely as possible the behavior in the real world.

One of the first experiments on preferences was performed in 1951 by M.M. Flood at the Rand Corporation\(^6\). Preferences were elicited by asking subjects to rank goods and money. Half of the subjects showed intransitive preferences.

In an experimental determination of indifference curves, MacCrimmon and Toda (1969) used an incentive to get the true preferences by selecting at random a point in the commodity space. If this point is below the indifference curve, then the subject will receive a commodity bundle on the curve. The goods were money and ball point

---

5 Using family budget data for 188 goods during 180 months (see Mattei (1994)), we obtained an average power of just 0.699. The power varies from household to household with a minimum of 0.370 and a maximum of 0.948.

6 This experiment is cited by May (1954) as another evidence of intransitive preferences. His experiment concerned the choice among three hypothetical marriage partners.
pens in the first session\textsuperscript{7}. Subjects had to draw seven indifference curves. None of the curves intersected but a request to select between pairs of commodity bundles showed inconsistent results. In the second experiment with money and French pastries, five students had to eat more than 4 pastries before getting the money. In this case the slopes of the indifference curves were positive and violation of transitivity was found for three subjects.

Battalio et al. (1973) used a token economy as a laboratory for the experimental analysis of economic behavior. Psychologists at the Central Islip State Mental Hospital in New York had developed a token economy for therapeutic reasons. Observations of consumer purchases of three groups of goods were obtained for seven weeks\textsuperscript{8}. During these weeks, Battalio et al. made large changes of relative prices in order to have a high number of budget set intersections (88\% of all possible intersections\textsuperscript{9}). The behavior of half of the 38 patients was inconsistent. The reporting errors in the data and the special environment of a psychiatric hospital diminish the importance of this interesting experiment.

An experiment with real consumption of the goods chosen was recently performed by Sippel (1997). Students of law or economics were paid 25 DM to participate in a laboratory experiment where they had to choose among eight goods in ten different budget situations. One of the ten bundles of goods was drawn at random and consumed on the spot.

In a first experiment, 12 subjects had to fill out 10 order sheets and then spend 60 minutes to consume the goods. The choices of eleven subjects out of twelve did not

\textsuperscript{7} This experiment was recently replicated by Knetisch (1992) with 20 women interviewed at a public swimming pool.

\textsuperscript{8} For a modern analysis of these data and same extensions, see Cox (1997) and Varian (1991).

\textsuperscript{9} The average power of the test, computed by Cox (1997, Table 5) is 0.702.
satisfy the strong axiom of revealed preference. Only 7 subjects (58.3%) satisfied the
generalized axiom of revealed preference.

In a second experiment with 30 students, Sippel implemented a Slutsky-type com-
pensation for the price changes. In this case, 22 subjects (73.3%) did not satisfy the
strong and 19 (63.3%) the generalized axiom of revealed preference.

While Sippel’s experiments are a great improvement over previous studies, they are
not as closely as possible to the behavior of a regular consumer. In particular, all the
goods had to be consumed in the laboratory during a limited period of time (1 hour).

In this paper I report the results of three experiments with a total of about 450
individuals. Real goods, real money and the same incentives as in the real world are
used.

The paper is organized as follows. Section II provides the theoretical framework of
the nonparametric tests which are used to check if the data are consistent with utility
maximization. In Section III the real power of these tests is evaluated using a model
of random choices. The experiments and their results are presented in Section IV. The
tests with a flexible function form are reported in Section V. Finally, Section VI offers
some concluding remarks.
II. The theoretical framework

The revealed preference theory of Samuelson (1938), extended by Houthakker (1950), can be presented in several ways. We will restate Varian’s version which is based on previous works of Koo (1963), Dobell (1965), Afriat (1967) and Warshall (1962), among others.

Let \( q^i = (q^i_1, q^i_2, \ldots, q^i_m) \) denote the vector of the quantities bought in period \( i \) and \( p^i = (p^i_1, p^i_2, \ldots, p^i_m) \) the vector of the corresponding prices. We have \( T \) observations of these prices and quantities \((p^i, q^i), i = 1, 2, \ldots, T\). The revealed preference theory tells us that:

1. \( q^i \) is directly revealed preferred to \( q^j \), written \( q^i R^o q^j \), if \( p^i \cdot q^i \geq p^j \cdot q^j \) where \( p^i \cdot q^j \) is the inner product of the vectors \( p^i \) and \( q^j \).

2. \( q^i \) is revealed preferred to \( q^j \), written \( q^i R q^j \), if there is some sequence \( q^i, q^a, q^b, \ldots, q^g \) such that \( p^i \cdot q^i \geq p^i \cdot q^a, p^a \cdot q^a \geq p^a \cdot q^b, \ldots, p^g \cdot q^g \geq p^g \cdot q^j \). The relation \( R \) is called the transitive closure of the relation \( R^o \).

A set of data satisfies the Generalized Axiom of Revealed Preference (GARP) if \( q^i R q^j \) implies \( p^i \cdot q^j \leq p^j \cdot q^i \).

This axiom allows multivalued demand functions whereas the more common Strong Axiom of Revealed Preference (SARP) [ \( q^i R q^j \) and \( q^i \neq q^j \) implies not \( q^i R q^i \) ] requires single valued demand function (See Varian (1982), p. 947).

Varian has shown that a set of data is consistent with the model of utility maximization if and only if it satisfies the Generalized Axiom of Revealed Preference.

We can check if the data are consistent with GARP by constructing a \( T \) by \( T \) matrix \( A \) whose element \( a_{ij} \) is equal to 1 if \( p^i \cdot q^i \geq p^i \cdot q^j \) (that is, if \( q^i R^o q^j \)) and 0 otherwise. The transitive closure \( R \) can be computed using the Warshall algorithm, modified by Varian (See his algorithm 5). This algorithm creates a matrix \( B \) whose element \( b_{ij} \) is equal to 1 if \( q^i R q^j \) and 0 otherwise. A violation of GARP is obtained if \( b_{ij} = 1 \) and \( p^i \cdot q^j > p^i \cdot q^i \).
When we use real data, we have to be careful in applying this test of revealed preference. Indeed, a difference of just one cent (of a franc or dollar) between the cost of two bundles is enough to reject the neoclassical model of consumer behavior\textsuperscript{10}. Moreover, a "nearly optimizing behavior" might be more appropriate for an individual consumer. Therefore, it is worthwhile to use a weaker test.

We can say that \( q^i \) is directly revealed preferred to \( q^j \) if the cost of \( q^i \) is significantly different from that of \( q^j \). Formally, \( q^i \ R^0 \ q^j \) if

\[
\epsilon \ p^i \cdot q^i \geq p^j \cdot q^j \quad 0 \leq \epsilon \leq 1
\]

where \( \epsilon \) is called the Afriat efficiency index\textsuperscript{11}. For example, if \( \epsilon = 0.8 \) then the cost of bundle \( q^i \) must be 25\% greater than that of bundle \( q^j \) in order to say that \( q^i \) is directly preferred to \( q^j \). A similar difference should be allowed for the violation of GARP.

We can check if the data are consistent with a "near optimizing behavior" by constructing a matrix \( A \) whose element \( a_{ij} \) is equal to 1 if \( \epsilon \ p^i \cdot q^i \geq p^j \cdot q^j \) and 0 otherwise. The transitive closure is computed as before, using the Warshall algorithm. A violation of a weakened GARP\textsuperscript{12} is obtained if \( b_{ij} = 1 \) and \( \epsilon \ p^i \cdot q^j > p^j \cdot q^i \).

\textsuperscript{10} A floating point error in the nonparametric test can lead to a wrong rejection of the model or fail to detect an inconsistent behavior. Cox (1997) found that Varian’s Nonpar software did not detect two violations in his data (See his footnote 11). We will use a program where the floating point errors are taken into account.

\textsuperscript{11} See Varian (1990), p. 130. It is possible to choose a different \( \epsilon \) for each observation or to search for the indices that are as closely as possible to 1. We have computed all these indices but we will report only the results with a single \( \epsilon \) (see Afriat (1967)) which is the most satisfactory statistic about the overall consistency of the data.

\textsuperscript{12} Or \( GARP_\epsilon \) (see Varian (1990)).
III. The power of the test

It is important to analyze the data before checking for consistency with preference maximization because if there are no budget set intersections, it is impossible to find violations of GARP. Every chosen bundle of goods which exhausts the budget is consistent with GARP. The power of the test is zero.

Bronars (1987) calculates the approximate power of the test by using random consumption data which exhaust the budget set. The data are obtained by using an algorithm which generates random budget shares. His second algorithm\textsuperscript{13} computes the budgets shares as follows:

$$\omega_i = \frac{z_i}{\sum z_i}$$

where $z_i$ are i.i.d. uniform random variables. Budget shares are multiplied by total expenditure and divided by the prices of the corresponding goods to obtain the random consumption data. The percentage of times that GARP is rejected gives the power of the test. Table I presents the power of the test when the prices and total expenditure are those of our second experiment (see below). These figures were obtained using 1000 random consumption data.

The first line of this table implies that GARP is not satisfied with random choices. The power of the test is very high. One can accept the model of utility maximization if the data pass this test. The hypothesis of a "near optimizing behavior" is much more difficult to test. If a waste of 5\% of the income is considered as an acceptable error, then it becomes very difficult to discriminate between a random choice and a utility-maximizing behavior (56.8\% of random choices are accepted as consistent choices).

\textsuperscript{13} Bronars uses three different algorithms to generate the random consumption data. The results obtained with all these algorithms are very similar.
Table I: The power of the test

<table>
<thead>
<tr>
<th>Afriat efficiency index</th>
<th>Percentage of inconsistent behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.994</td>
</tr>
<tr>
<td>0.99</td>
<td>0.952</td>
</tr>
<tr>
<td>0.98</td>
<td>0.886</td>
</tr>
<tr>
<td>0.97</td>
<td>0.777</td>
</tr>
<tr>
<td>0.96</td>
<td>0.594</td>
</tr>
<tr>
<td>0.95</td>
<td>0.432</td>
</tr>
<tr>
<td>0.90</td>
<td>0.029</td>
</tr>
</tbody>
</table>

A low value of this approximate test means that it is very unlikely to find violations of GARP.

To check if it is impossible to find violations of GARP, a second test can be used. Since a budget set intersection is a necessary condition to find a violation of GARP, one can estimate the power of the test by calculating the number of budget set intersections. We have a budget set intersection between period $t$ and period $\tau$ if for some good $i$ $\frac{y_t}{p_{ijt}} > \frac{y_{\tau}}{p_{i\tau}}$ and for some other good $j$ $\frac{y_t}{p_{jyt}} < \frac{y_{\tau}}{p_{j\tau}}$ (or vice versa) where $y_t$ is total expenditure in period $t$. It is useful to express budget set intersections as a percentage of the maximum number of these $2 \times 2$ comparisons which is $\left(\begin{smallmatrix} T \\ 2 \end{smallmatrix}\right)$.

When the prices and total expenditure are those of our second experiment, we obtained 189 intersections out of a total of 190 (if you want to test homogeneity you cannot have a budget set intersection in at least one case).
IV. The experiments

In a first experiment, I recruited 20 students attending my microeconomics class. The theory of revealed preference was one of the topics of the course. These students had to choose among eight goods in twenty different budget situations. I told them that I wanted to simulate the behavior of a consumer going to 20 different stores with a given amount of money to spend in each store. As an incentive to participate in the experiment and to behave as a real consumer, I inform them that they will receive one of these bundles of goods.

The amount of money to spend in each store varied between 30 and 40 Swiss francs. This means that the value of the chosen bundle of goods was about 35 Swiss francs (23 $).

We choose eight goods that could interest the students and were not too difficult to distribute (see Appendix 5).

The instructions for the experiment were then distributed (see Appendix 1 for similar instructions used in the third experiment\(^{14}\)). To participate in the experiment, the students had to go to the computer center and use a program based on the EXCEL spreadsheet. To facilitate the computations, we prepared a program, written in VISUAL BASIC, with 20 spreadsheets corresponding to the different stores to visit. The subjects had to fill in the quantities bought in each store. All the informations were available on the screen, in particular the prices of each good.

In the first store, the prices were those of a grocery store. The prices in the other stores were increased or decreased in order to have a maximum number of budget intersections.

The program checked and informed the subject if the budget constraint was satisfied or not and the amount left, if any. After the quantities for the 20 stores were chosen, the program recorded in a file all the choices of the subject and his name.

\(^{14}\) The original instructions were in French.
Table II gives the results of the test. With 25% of inconsistent subjects, we have a better result than previous experiments. Taking a course in microeconomics may improve the behavior of the subjects. Nevertheless, it would be wrong to assume, without qualifications, that the neoclassical model describes the behavior of every consumer. A possibility to solve this problem is to accept some kind of errors on the part of the consumer. Indeed, with an efficiency index of 0.95 all the subjects are consistent but the power of the test is just 0.432. Hence, it is premature to draw a conclusion for these five subjects.

Without transaction costs, the best behavior is to buy in each store only the relatively cheapest good and then exchange or sell it to get the desired quantities. We did not observe such a behavior but after the bundles of goods were distributed, some participants exchanged their goods.

The subjects of our second experiment were business students. We wanted to investigate if our results were robust by taking a large number of participants. The instructions for the experiment were distributed to about 500 students and 100 subjects participated in this second experiment. As with the first experiment, they had to go to the computer center and use a program based on the EXCEL spreadsheet. Half of the goods used in the first experiment and the prices were changed (see Appendix 6).

In spite of the warning from the computer, 4 subjects exceeded the budget constraint by more than 1%. On the other hand, 55 subjects did not use more than 1% of the amount available in at least one store\textsuperscript{15}. The expected value of this residual amount is just 2.5 cents. Hence, the incentive to use the available amount to the last penny is not very high. Indeed, some students did not bother to spend much time for such small amounts.

\textsuperscript{15} Six subjects did not use more than 10% of the available amount (in at least one store) and one subject exceeded the budget constraint by more than 10%. Three subjects did not use more than 5% of the available amount.
### Table II: Experiment I

<table>
<thead>
<tr>
<th>Afriat efficiency index</th>
<th>Number of inconsistent subjects</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>5</td>
<td>25%</td>
</tr>
<tr>
<td>0.99</td>
<td>2</td>
<td>10%</td>
</tr>
<tr>
<td>0.98</td>
<td>1</td>
<td>5%</td>
</tr>
<tr>
<td>0.95</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

### Table III: Experiment II

<table>
<thead>
<tr>
<th>Afriat efficiency index</th>
<th>Number of inconsistent subjects</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>44</td>
<td>44%</td>
</tr>
<tr>
<td>0.99</td>
<td>30</td>
<td>30%</td>
</tr>
<tr>
<td>0.98</td>
<td>16</td>
<td>16%</td>
</tr>
<tr>
<td>0.95</td>
<td>4</td>
<td>4%</td>
</tr>
</tbody>
</table>

### Table IV: Experiment III

<table>
<thead>
<tr>
<th>Afriat efficiency index</th>
<th>Number of inconsistent subjects</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>101</td>
<td>32%</td>
</tr>
<tr>
<td>0.99</td>
<td>66</td>
<td>21%</td>
</tr>
<tr>
<td>0.98</td>
<td>50</td>
<td>16%</td>
</tr>
<tr>
<td>0.95</td>
<td>7</td>
<td>2%</td>
</tr>
</tbody>
</table>
We would have encountered the same problems even if all the goods chosen were distributed. There is always a last penny to be spent. Moreover, we do not know the motivations of the participants. Several experiments with the ultimatum bargaining game show that an increase in the financial stakes does not alter the conclusions (see Slonim and Roth (1998)). Some participants complained that the experiment was too long. We found that only 16 subjects did not use more than 1% of the amount available in the first store. Actually, there is a decrease in the motivation to use all the available amount.

If satiety is excluded, the residual amount should not exceed the price of the cheapest good. This is the case for half of the participants. However, for most of the goods fractional quantities were allowed in order to arrive as close as possible to the available amount (see Appendix 2 for a similar case). The residual amounts of 14 students satisfied this condition.

We think that it would be wrong to enforce a strict observance of the budget constraint. The consumers have only a rough idea of the budget constraint. They often pay their purchases with a credit card. Normally, they do not spend all the money that they have in their purse. Moreover, it is very easy to give an example of an inconsistent behavior as a consequence of a strict observance of the budget constraint.

A tourist buys a bottle of Coca-Cola and two sandwiches when their prices are, respectively, 3 and 7 francs. She then arrives at the airport with 14 francs of coins. She wants to spend all this amount because the coins are worthless outside the country. The prices of the two goods are 5 and 4 francs, respectively. In order to spend 14 francs she has to buy 2 bottles of Coca-Cola and a sandwich. This yields an inconsistent behavior. If we require that all the money must be spent, we force the consumer to have an inconsistent behavior. A consistent behavior would be to continue to buy a bottle of Coca-Cola and two sandwiches even if not all the coins are used.
Hence, we think that the best solution it to take the amount spent in each store as the effective budget constraint of the participant. Indeed, the inconsistent behaviors are greater among the subjects that strove to seek the combination of goods which satisfy the budget constraint (57% of inconsistent subjects among the 14 participants with strict observance of the budget constraint\(^{16}\)). A similar result will be reported below.

The changes in the power of the test due to the uses of the amount spent as the new budget constraint were insignificant. The average power was 0.989 for \(\epsilon=1\) (instead of 0.994) and 0.431 for \(\epsilon=0.95\) (instead of 0.432).

Table III gives the results. The inconsistent subjects were 44 and a difference of 5% is not enough to account for a "near optimizing behavior".

In this second experiment, we also wanted to test if the individual demands were homogeneous of degree zero. The prices of store 17 and the amount of money available there were 25% higher than those of store 12. This test requires that the consumer spends exactly all the available amount of money. Only in 4 of the 39 cases where this condition was satisfied, the hypothesis is accepted\(^ {17}\). The average angle between the consumption vectors in stores 12 and 17 is 20 degrees. We have here a striking example where a hypothesis may be valid for a representative consumer but is clearly not fulfilled for the individual consumer.

To facilitate the distribution of the rewards and at the same time to test the superiority of money over goods, we let the participants choose between 50 Swiss francs and the bundle of goods that they selected in one of the store drawn at random. We wrote a letter asking the participants if they wanted to buy themselves the bundle of goods (see Appendix 4 for a similar letter used in the third experiment). As expected, all the subjects preferred to receive 50 francs even if in 16 cases the goods were worth more than 55 francs. This result does not mean that the subjects did not maximize their

\(^{16}\) A statistical test of the difference between the two proportions gives a p-value of 0.36.

\(^{17}\) The behavior of 41 subjects was consistent with SARP.
utility because the eight goods are only a very small part of all available goods. This result is very robust and confirms a classical statement found in every textbook: to give money is always better than to give goods as in the U.S. food stamp program.

The great majority of subjects used in experimental economics are students just because it is easier to recruit them. However, the behavior of a student may not correspond to that of the vast majority of consumers\textsuperscript{18}. To analyze the behavior of experienced consumers we put an announcement in a magazine for consumer affairs (40000 circulation copies). We had 434 consumers interested in our experiment. We sent the instructions to these consumers (see Appendix 1). In this experiment it was not possible to use our computer program. Hence, we prepared a questionnaire where the participants had to write their choices (see Appendix 2). A table with all the prices and the amounts available were enclosed so that one can see at once all the relevant data as in a perfect information case (see Appendix 3). Moreover, in each store the list of goods whose prices have increased or decreased with respect to those of the previous store was also given. We changed some of the goods to take into account the tastes of these consumers. The prices and the available amount of money were those of our second experiment.

Without a computer, it is more difficult to satisfy the budget constraint but we expected that the participants were used to encounter similar problems with their consumption expenditure.

\textsuperscript{18} In particular, economics students might adapt their behavior to the axioms of the theory they study (See Carter and Irons (1991)).
The returned questionnaires were 320. As expected, the budget constraint was more easily exceeded in this experiment. In 4% of the subjects the difference was in some cases beyond 10% of the amount available. On the other side, 12% of the participants did not use more than 10% of the amount available in at least one store\textsuperscript{19}.

Only 2 subjects did not use more than 10% of the amount available in the first store. As in the previous experiment, we found a decrease in the motivation to use all the available amount.

If satiety is excluded, the residual amount should not exceed the price of the cheapest good. This is the case for 90 participants. However, for most of the goods fractional quantities were allowed in order to arrive as close as possible to the available amount (see Appendix 2). The residual amounts of 40 students satisfied this condition.

From the numerous comments received, we are convinced that the participants did their best to satisfy the budget constraint. Of course, we could have helped them with a computer program but our aim was to simulate the real behavior of the consumers. We did not want to interfere in their choices. Moreover, we wanted to check if there was a difference with respect to the second experiment.

For the same reasons as those given for the second experiment, we take the amount spent in each store as the effective budget constraint of the participants. The changes in the power of the test due to the uses of the amount spent as the new budget constraint were insignificant. The average power was 0.989 for $\epsilon=1$ (instead of 0.994) and 0.428 for $\epsilon=0.95$ (instead of 0.432).

Table IV gives the results of the test. The inconsistent subjects were 101 (32%). This percentage is between the value obtained in the first and that of the second experiment.

\textsuperscript{19} In 20% of the subjects the difference was beyond 1% of the available amount (in at least one store) and 65% of the participants did not use more than 1% of the available amount. The percentages for a residual amount greater than 5% (in absolute value) are, respectively, 7% and 19%.
The uniformity of the results of these and other experiments is striking. A significant number of subjects have an inconsistent behavior. Therefore, we can conclude that it is wrong to assume that all individual consumers have a utility maximizing behavior.

With an efficiency index of 0.95, only 2% of the subjects have an inconsistent behavior. However, as Table I shows, the power of the test is not very high in this case.

As in the previous experiment, the percentage of inconsistent subjects (56.3%) is greater among the 32 consumers with a strict observance on the budget constraint.

The robustness of our results can also be tested with a sensitivity analysis. We perturbed the expenditure on each good using 1000 i.i.d. uniform random variables with a mean deviation of 0.50 francs. Only 20.6% of the inconsistent subjects become consistent whereas 90.1% of the consistent subjects remain consistent.

As in the second experiment, we could test if the individual demands were homogeneous of degree zero. The prices of store 17 and the amount of money available there were 25% higher than those of store 12. This test requires that the consumer spends exactly all the available amount of money. Only in 4 of the 116 cases where this condition was satisfied, the hypothesis is accepted\textsuperscript{20}. The average angle between the consumption vectors in stores 12 and 17 is 20 degrees.

To facilitate the distribution of the rewards and at the same time to test the superiority of money over goods, we let the participants choose between 50 Swiss francs and the bundle of goods that they selected in one of the store drawn at random.

We wrote a letter asking the participants if they wanted to buy themselves the bundle of goods (see Appendix 4). As expected, 96.6% of the subjects preferred to receive 50 francs even if for 32% of these participants the goods were worth more than 55 francs. Ten of the remaining subjects eventually accepted to receive 50 francs. The reason given to prefer the goods was that they just wanted to play the experiment to the end.

\textsuperscript{20} The behavior of 153 subjects (48\%) was consistent with SARP.
If we give a normative interpretation to the neoclassical model then we could say that the inconsistent consumers should be ready to correct their errors. We did not test this hypothesis because we expected that the majority of these consumers would answer that they are indifferent between their choices and the corrected bundles of goods.

A more promising solution of this problem is the rehabilitation of the hypothesis of the representative consumer.

The neoclassical model of consumer behavior was never intended to describe the behavior of "an actual person, the Mr. Brown or Mr. Jones who lives round the corner" (Hicks (1956, p. 55)). The demand function of Mr. Brown may be very special or idiosyncratic and without any analytical interest.

The difference between economics and psychology is that economics is not interested in the behavior of an individual consumer per se but only inasmuch as he shapes the behavior of a significant group of individuals. As pointed out by Marshall: "There are many classes of things the need for which on the part of any individual is inconstant, fitful, and irregular ... But the economist has little concern with particular incidents in the lives of individuals. He studies rather "the course of action that may be expected under certain conditions for the members of an individual group" ... and in these broad results the variety and the fickleness of individual action are merged in the comparatively regular aggregate of the action of many." (Marshall (1920), p. 83). Cournot has a similar statement: "Nous admettrons que la fonction F(p) qui exprime la loi de la demande ou du débit est une fonction continue ... Il en pourrait être autrement si le nombre des consommateurs était très limité: ainsi, dans tel ménage, on pourra consommer précisément la même quantité de bois de chauffage, que le bois soit à 10 francs ou à 15 francs le stère; et l'on pourra réduire brusquement la consommation d'une quantité notable, si le prix du stère vient à dépasser cette dernière somme. Mais plus le marché s'étendra, plus les combinaisons de besoins, des fortunes ou même des caprices, seront
variées parmi les consommateurs, plus la fonction $F(p)$ approchera de varier avec $p$
d'une manière continue". (Cournot (1838, ch. IV))

We will define the representative consumer as the hypothetical individual whose be-
havior corresponds to the average behavior of a group of consumers with the same
income and prices. The heterogeneity of preferences should make the aggregate behav-
ior more regular (see Grandmont (1987)) and the individual errors should be cancelled
out on average.

The data of our three experiments confirm this hypothesis. The behaviors of all these
three average consumers are consistent.

Appendix 7 gives the mean budget shares of the representative consumer, their stan-
dard deviations and the highest standard deviations for the individual participants. As
expected, the variations of the budget shares of the representative consumer are more
regular and less abrupt.

In the next section, I will use these data to estimate the demand functions and to
test the symmetry and homogeneity restrictions.
IV. A parametric test using a flexible functional form

The demand functions represent the most important materialization of consumer preferences. The neoclassical model imposes several restrictions on these functions. Since the demand functions are an indispensable tool in applied economics, it is crucial to check if the theoretical restrictions are satisfied. This was the approach taken by the first tests of consumer theory.

A parametric estimation is obtained by considering the data as a sample and therefore an error term is always added to the function to be estimated. This term can incorporate all kind of errors due to missing variables or any other reason. A trembling-hand type of behavior can be one of these reasons. For instance, take the case of the linear expenditure system. The demand function is:

\[ q_i = c_i + \frac{b_i}{p_i} (R - \sum p_j c_j) \]

where \( c_i \) and \( b_i \) (\( \sum b_i = 1 \)) are parameters to be estimated and \( R \) is disposable income (total expenditure). If we add an error term with mean zero and a given standard deviation, then the values taken by the \( q_i \) can be inconsistent with the theory of revealed preference in spite of a well-behaved model. Assume that the vectors of \( c_i \) and \( b_i \) are, respectively, \([0.1, 0.2, 0.3, 0.1, 0.2, 0.1, 0.2, 0.3]\) and \([0.1, 0.1, 0.2, 0.1, 0.2, 0.1, 0.1, 0.1, 0.1]\). We can generate the quantities by using the prices and income of our second experiment. When the error term is a normal random variable with mean zero and standard deviation of 0.64, then the obtained quantities are inconsistent with the theory of revealed preference. With an efficiency index of 0.99 no inconsistent behavior is detected. Hence, the results obtained with the efficiency index can be compared to the tests with a parametric model.

Since our data for the representative consumer are consistent with utility maximization, it is interesting to compare the results obtained with a parametric test.
Deaton and Muellbauer (1980a) have proposed a model (called the Almost Ideal Demand System) which gives a local first-order approximation to any demand system. This system of demand equations has been obtained by using a class of preferences which permit exact aggregation over consumers. The demand functions in budget share form are:

\[ w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log \{R/P\} \]  

(1)

where

\( w_i = \) budget share of good \( i \) \( (p_i q_i / R) \)

\( p_i = \) price of good \( i \)

\( R = \) total expenditure

\( P = \) a price index defined by

\[ \log P = \alpha_o + \sum \alpha_k \log p_k + 0.5 \sum_j \sum_k \gamma_{kj} \log p_k \log p_j \]

The adding-up property of the budget shares \( \sum w_i = 1 \) implies the following restrictions on the parameters:

\[ \sum \alpha_i = 1 \; ; \; \sum_i \gamma_{ij} = 0 \; ; \; \sum \beta_i = 0 \]

Homogeneity of degree zero in prices and total expenditure requires:

\[ \sum_j \gamma_{ij} = 0 \]

and the Slutsky symmetry condition is satisfied if

\[ \gamma_{ij} = \gamma_{ji} \]

If \( P \) is approximated by \( P^\ast (P \approx \Phi P^\ast) \) where
\[ \log P^* = \sum w_k \log p_k \]

then the demand functions become:

\[ w_i = \alpha_i^* + \sum_j \gamma_{ij} \log p_j + \beta_i \log \{R/P^*\} \tag{2} \]

where

\[ \alpha_i^* = \alpha_i - \beta_i \log \Phi. \]

This equation can be estimated for each good separately by using ordinary least squares. Given normally distributed errors, ordinary least squares estimation is equivalent to maximum likelihood estimation for the whole system.

Table V reports the parameter estimates obtained with the data of the representative consumer in our third experiment. Twenty-six coefficients \( \gamma_{ij} \) out of sixty-four have \( t \)-values larger than 2.

The column headed \( \sum \gamma_{ij} \) gives the row sums of these coefficients. These sums should be zero if homogeneity holds. A \( t \)-test can be used to check if the sum is statistically different from zero. The 5 percent significance limit for 10 degrees of freedom is 2.228. Hence, we have to reject this hypothesis for the demand of biscuits.

If the homogeneity restriction is imposed, we can estimate:

\[ w_i = \alpha_i^* + \sum_{j=1}^{m-1} \gamma_{ij} \log (p_j/p_m) + \beta_i \log \left( R/P^* \right) \tag{3} \]

The last three columns of Table V give equations standard errors, the \( R^2 \) and Durbin-Watson statistics for equations (2) and (3).

Instead of a \( t \) test, we could have tested the homogeneity restrictions by considering \( F \)-ratios of the sum of squares of restricted and unrestricted residuals (the \( F \)-ratios are the squares of the \( t \)-values).
The last column shows that the imposition of homogeneity does not lead to a fall in the Durbin-Watson statistics when this restriction is rejected. The results obtained with time series data do not hold in this case of random values.

To test symmetry, we must estimate simultaneously all the demand equations since we have restrictions across equations. Given the high number of parameters in each equation, Deaton and Muellbauer assume a particular structure of the variance-covariance matrix of the residuals (an identity matrix modified to take into account the singularity of the variance-covariance matrix of the residuals due to the adding-up restriction) so that a maximum-likelihood estimation reduces to least squares.

The system as a whole was estimated using a nonlinear least squares program written by Bard (1974, pp. 96-106). All the restrictions are imposed by substitution. We take the 42 independent parameters (28 for the \( \gamma_{ij} \), 7 for the \( \alpha_i \) and the \( \beta_i \)) and use the restrictions to determine the 10 other parameters. To test the validity of the restrictions, we employ the likelihood ratio \( \lambda \).

Table VI reports the values of 2 log likelihood using \( P^* \) and \( P \). The likelihood ratio statistic for symmetry given homogeneity is 25.15 (\( \chi^2_{21,0.01}=38.9 \)); for symmetry and homogeneity jointly it is 47.42 (\( \chi^2_{28,0.01}=48.3 \))\(^{21}\). These restrictions are accepted at the 1% signification level. Hence, there is no basic contradiction between the nonparametric test and the parametric test. This conclusion is even more valid if we take into account the approximations of the asymptotic tests.

\(^{21}\) The likelihood ratio statistic for joint homogeneity is 22.27 (\( \chi^2_{7,0.01}=18.5 \)). As it is well known, this test is biased toward rejecting the restrictions. Laitinen (1978) has shown that the Wald statistics is distributed as Hotelling's \( T^2 \). In our case, the Wald statistics for joint homogeneity is 240.89. The corresponding \( T^2 \) 1% critical value is 262.15.
Table VI

Almost Ideal Demand System

Comparative values of 2 Log Likelihood

<table>
<thead>
<tr>
<th>Constraints</th>
<th>Using P*</th>
<th>Using P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted</td>
<td>-</td>
<td>950.64</td>
</tr>
<tr>
<td>Homogeneous</td>
<td>7</td>
<td>928.99</td>
</tr>
<tr>
<td>Symmetric</td>
<td>21</td>
<td>-</td>
</tr>
</tbody>
</table>
VI. Concluding remarks

This paper shows that the behavior of a significant number of individual consumers is inconsistent with the neoclassical model, at least in its textbook or naïve form. The results of these and other experiments cast serious doubts on the careless uses of this model. However, the neoclassical model was never intended to describe the behavior of an individual consumer. The consumer of the theory is an ideal individual whose behavior should correspond to that of a group of consumers with the same income and confronted with the same prices.

The results of our three experiments confirm this hypothesis. The data for the average or representative consumer support the neoclassical model of consumer behavior.

A parametric test with these average data gives the same result. If we take into account that the functional form and the test are approximations then we can conclude that the hypothesis of homogeneity and symmetry are accepted.

It would be interesting to check if the intertemporal version of the neoclassical model withstands the same tests.
References

Afriat, S.N. (1967), ”The Construction of a Utility Function from Expenditure Data”,


Battalio et al. (1973) ”A Test of Consumer Demand using Observations of Individual

Economy, vol. 70, 1962, pp. 1-13


Carter, J.R., and M.D. Irons, ”Are Economists Different, and If So, Why?”, Journal of

Christensen, L.R. D.W. Jorgenson, and L.J. Lau (1975), ”Trascendental Logarithmic

Cournot, A. (1838), Recherches sur les principes mathématiques de la théorie des
richesses, Hachette, Paris, 1838

1997, pp. 1054-1078

University Press, Cambridge, 1980

Deaton, A. and J. Muellbauer (1980a), ”An Almost Ideal Demand System”, American

Dobell, A.R. (1965), ”A Comment on A.Y.C. Koo’s ”An Empirical Test of Revealed

Grandmont, J.-M. (1987), ”Distributions of Preferences and the ”Law of Demand” ”,


Appendix 1

Instructions for participants

The purpose of this experiment is to analyze the behavior of consumers when prices change. We want to simulate the purchases done by an individual with a certain amount of money at his disposal.

The goods that you can buy are real goods that you will find in any local store. You will receive these goods as compensation for your participation in the experiment.

You have at your disposal a given amount of money which varies between 42 and 56 francs depending on the store. With this money you have to buy goods whose prices change from store to store. When the experiment is ended, you will receive the goods that you bought in one of the stores (selected at random). Therefore, by participating to this experiment, you can get goods whose value is about 50 francs.

In the enclosed questionnaire you will find 20 tables with the prices of the goods in different stores. We suggest to write in pencil the quantities that you want to buy. In this way, you can easily make changes if you exceed the amount of money at your disposal or if there is some money left.

You obtain the expenditure by multiplying the quantity by the price of the good in the store. Beware that total expenditure cannot exceed the amount that you have for the purchases in the store.

In page 1 you will find the data for the first store. You have the list of the goods and their prices. The column "base prices" gives the actual prices of these goods. These prices correspond to the amounts that you have to pay for these goods in a grocery store. The column "store’s price" gives the prices in the store. For store number 1 these prices are the same as the "base prices".

The last line tells you that you have at your disposal 50 francs. You have to spend this amount in this store. You cannot use this money to buy goods in other stores. You will have other amounts at your disposal for the next stores.