Utility Maximization

James C. Cox
University of Arizona

The model of the utility-maximizing agent is central to theoretical and applied economics and to economics education. But how can one test for the presence of utility maximization? More specifically, how can one test “the utility hypothesis,” defined as the hypothesis that real economic agents behave as if they maximize utility functions subject to binding budget constraints?

1. The Utility Hypothesis

A testable form of the utility hypothesis results from identifying necessary condition(s) for the existence of a utility function that “rationalizes” the price and quantity data, meaning a utility function that correctly predicts choices that do not violate an agent’s budget constraint. Consider the traditional model of a “consumer” that chooses only consumption good quantities. Let the data consist of price vectors, p^t, and quantity vectors, q^t, for observations, $t = 1,2,\ldots,T$. The consumer has been observed to choose q^t, which therefore did not violate her budget constraint. The cost of q^t was $p^t \cdot q^t$; hence any q such that

(1) $$p^t \cdot q \leq p^t \cdot q^t$$

would also not violate the situation t budget constraint. Thus the utility function $u(\cdot)$ rationalizes the (p^t, q^t) observations if and only if $u(q^t) \geq u(q)$ for all q such that $p^t \cdot q^t \geq p^t \cdot q$. Thus we have:

Statement 1: If there exists a utility function, $u(\cdot)$*, that rationalizes data for a consumer then that agent behaves as if he maximizes* $u(\cdot)$*; that is, his behavior is consistent with the utility hypothesis*.

Possible inconsistencies with utility-maximizing choices are of no clear economic interest unless the objects of choice are scarce. In order for the chosen commodities to be scarce, the agent’s budget constraint must be binding. Hence, a specific testable form of the utility hypothesis must incorporate scarcity. There are various ways to incorporate scarcity in the model of the utility-maximizing agent. (Sufficient conditions include assumptions that the agent's preferences are: (a) strictly monotonic; or (b) globally non-satiated and strictly convex; or (c) locally non-satiated.) I shall assume that the agent's preferences are locally non-satiated, meaning that for any admissible commodity vector, q, and any neighborhood around q, there exists an admissible q^* in the neighborhood that is strictly preferred to q.

A key theoretical result is derived in Varian (1982). He demonstrated that the necessary and sufficient condition for the existence of a locally non-satiated utility function that rationalizes consumption good price and quantity observations is the generalized axiom of revealed preference (*GARP*).

$$GARP\text{: } if\quad p^i \cdot q^i \geq p^i \cdot q^j,\, p^j \cdot q^j \geq p^j \cdot q^r,\ldots, p^s \cdot q^s \geq p^s \cdot q^k$$

$$then\quad p^k \cdot q^k \leq p^k \cdot q^i$$

Thus, in principle, a test of the utility hypothesis can be based on *GARP*.

But how, in practice, can one actually test the utility hypothesis with data for real economic agents? The utility hypothesis has no testable implication for aggregate data (Sonnenschein, 1973a,b) nor for proper subsets of the variables ("incomplete" data) that enter an agent's budget constraint (Varian, 1988). The utility hypothesis and the hypothesis that preferences are homogenous across consumers together place testable restrictions on aggregate data (Shafer and Sonnenschein, 1982). Therefore, if one conducts a test with aggregate data he is doing a joint test of a compound hypothesis consisting of the (core) utility hypothesis and the (subsidiary) hypothesis that preferences are homogenous across agents. The utility hypothesis and the hypothesis that preferences are weakly separable together place testable restrictions (Varian, 1983) on both the incomplete data (one set of inequalities) and the complete data (another set of inequalities). Therefore, if one conducts a test with incomplete data, he is doing an incomplete (one of two sets of inequalities) joint test of the compound hypothesis consisting of the utility and separability hypotheses. But the homogeneity and separability hypotheses are not central to economics while the utility hypothesis is central. The key test is one that is capable of yielding results that would be known to be inconsistent with the utility hypothesis: a "simple" test of the core hypothesis, not a (complete or incomplete) joint test involving one or more subsidiary hypotheses. Thus we have:

Statement 2. In order to conduct a simple test of the utility hypothesis, one must use data that are "disaggregated" (i.e., are for individual decision-makers) and "complete" (i.e., include observations of all of the choice variable that enter budget constraints).

The choice variables that enter the budget constraints of economic agents in national economies typically include labor supplies and asset net demands in addition to consumption good demands. Thus we have:

Statement 3. In most economies, a data set must include observations of consumption good demands, labor supplies, and asset (net) demands in order for it to be complete.

This statement implies that we need to identify necessary condition(s) for utility-maximizing choice of consumption goods, labor supplies, and assets. Cox (1997) reports the extended form of *GARP* that includes labor supplies and asset net demands. Let the asset price vector observed in situation *t* be v^t and the observed asset quantity vector be a^t. Let the observed wage rate vector be w^t and the observed labor supply vector be h^t. Let I^t be the maximum possible labor income that the agent can earn in situation *t*. The agent spends $p^t \cdot q^t$ of I^t on consumption goods and "spends" $I^t - w^t \cdot h^t$ of I^t on leisure. The consumption, labor supply, and asset net demand quantities, (q^t, h^t, a^t), were observed to have been chosen in situation *t*; hence they are known to have been affordable. Thus any (q, h, a) such that

(2) $$p^t \cdot q + I^t - w^t \cdot h + v^t \cdot a \leq p^t \cdot q^t + I^t - w^t \cdot h^t + v^t \cdot a^t$$

is also known to have been affordable in situation *t*. The utility function, *u(q,h,a)* rationalizes the observations if $u(q^t, h^t, a^t) \geq u(q,h,a)$, for all *(q,h,a)* that satisfy (2), for all observations *t=1,2,…T*. Hence the necessary and sufficient condition for the existence of a such a utility function is the following extended form of the generalized axiom of revealed preference that incorporates labor supply and asset (net) demand.

$$
\begin{aligned}
GARP_{LA}: \ if \ \ & p^i \cdot q^i + I^i - w^i \cdot h^i + v^i \cdot a^i \geq p^i \cdot q^j + I^i - w^i \cdot h^j + v^i \cdot a^j, \\
& p^j \cdot q^j + I^j - w^j \cdot h^j + v^j \cdot a^j \geq p^j \cdot q^r + I^j - w^j \cdot h^r + v^j \cdot a^r, \\
& \ldots, \ p^s \cdot q^s + I^s - w^s \cdot h^s + v^s \cdot a^s \geq p^s \cdot q^k + I^s - w^s \cdot h^k + v^s \cdot a^k, \\
then \ \ & p^k \cdot q^k + I^k - w^k \cdot h^k + v^k \cdot a^k \leq p^k \cdot q^i + I^k - w^k \cdot h^i + v^k \cdot a^i
\end{aligned}
$$

Thus we have:

Statement 4: With data from most economies, conducting a simple test of the utility hypothesis is equivalent to testing $GARP_{LA}$ *with data for individual economic agents.*

2. A Complete, Disaggregated Data Set

Since conducting a simple test of the utility hypothesis requires the use of data that are disaggregated and complete, one must attempt to find such data. It appears that the only disaggregated and complete data that exist are the data from the token economy experiment reported in Battalio, et al. (1973) and Kagel, et al. (1972). The experiment was conducted with 38 female patients that were part of the ongoing token economy in the Central Islip State Hospital. This was an essentially closed economy in which residents could earn tokens by performing various tasks. The tokens could be used to purchase consumption goods available within the hospital economy and could be carried over for indefinite periods. Thus the choice variables in the token economy were consumption good quantities, labor hours for various tasks, and token money balance.

The economics experimenters obtained permission to conduct an experiment with the token economy that involved introducing the large changes in relative prices of three groups of consumption goods reported in Table 1. The experiment was conducted over a seven week period. The data consist of individual subjects' weekly earnings, weekly expenditures on each of the three groups of consumption goods, and end-of-week token balances. Although the labor hours supplied to various tasks were not recorded, one can still test $GARP_{LA}$ because nominal wage rates were not changed; that is, since $w^t = w$, for all *t*, one can substitute week *t* labor income, y^t, for $w^t \cdot h^t$ and $w^s \cdot h^t$ in all revealed preference statements. Furthermore, since the one asset in the experiment (tokens) was the numeraire, its price was always equal to one and therefore one can substitute the token money balance at the end of week *t*, m^t, for $v^t \cdot a^t$ and $v^s \cdot a^t$ in all revealed preference statements.

3. Test Results and Power

Figure 1 (right scale) shows the number of observed violations of $GARP_{LA}$ by each of the 38 subjects reported in Cox (1997). (No results are reported for subject 3 because she left the token economy shortly after the economics experiment began.) Data for 24 out of the 38 subjects contain no observed inconsistencies with $GARP_{LA}$. Data for the other 14 subjects contain 54 observed inconsistencies with $GARP_{LA}$, with subject 29 accounting for 17 of them.

Counting the number of inconsistencies with revealed preference inequalities is not informative unless one can assess the power of the test. For example, if none of the budget hyperplanes intersected then we would necessarily count zero inconsistencies with all revealed preference inequalities but the test would be completely uninformative. Figure 1 shows a measure of the power of the test for the data for each subject that is reported in Cox (1997). The test power is measured with a randomization procedure based on Becker's (1962) random choice model. First, a computer algorithm generates all possible combinations of budget shares for the *n* commodities that sum to 1 on a 0.01 unit of divisibility. Then a sample of 1,000 random budget shares is drawn (with replacement) from a discrete uniform distribution on the generated set of possible budget shares for each subject for each data observation week. Let $\tilde{S}_j^t$, $j = 1, 2, ..., n$, be one element of the sample of budget shares drawn for a subject; then the random quantities, $\tilde{q}_j^t$, are calculated as

(3) $$\tilde{q}_j^t = \tilde{S}_j^t \; p^t \cdot q^t / p_j^t$$

where p_j^t and $p^t \cdot q^t$ are in the actual data. The random quantities, $\tilde{q}_j^t$, and actual prices, p_j^t, are then used as data in tests of $GARP_{LA}$. The frequencies with which this random choice model produces one or more violations of the relevant revealed preference inequalities are reported in Figure 1 (left scale).

For the 24 subjects with zero violations using actual data, the power measure has an average value of .507 and it varies from a low of .371 for subject 38 to a high of .664 for subject 26. Therefore, the absence of violations in the data for 24 subjects cannot be attributed to nonintersecting budget hyperplanes. The power measure for the 14 subjects with one or more violations has an average value of .474 and it varies from a low of .243 for subject 1 to a high of .591 for subject 25. It seems clear that the different test results for the 24 subjects with no violations and the 14 subjects with violations cannot be attributed to lower test power for the no-violations group. This conclusion is supported by the *negative* rank correlation of -.034 between the number of observed violations and the power measure.

4. Are the Inconsistencies with Utility Maximization Significant?

Simply counting the 54 violations of $GARP_{LA}$ committed by 14 out of the 38 subjects does not inform us about the significance of these inconsistencies with utility maximization. Do these 14 subjects “almost maximize” utility functions or are their choices “far from” utility-maximizing choices? A nonparametric approach to assessing the significance of deviations from utility maximization was introduced by Afriat (1972) and further developed by Afriat (1987) and Varian (1993).

This approach can be most easily explained using the conventional revealed preference model in which consumption goods are the only choice variables (Cox, 1997). Let q^t be the observed *n*-vector of commodities chosen in situation *t* and let p^t be the corresponding observed price vector. Varian's (1993) efficiency index approach can be explained as follows. For some numbers $e^t \in [0,1]$, redefine *GARP* as

$$GARP(e^t): if \;\; e^i p^i \cdot q^i \geq p^i \cdot q^j, e^j p^j \cdot q^j \geq p^j \cdot q^r, \ldots, e^s p^s \cdot q^s \geq p^s \cdot q^k$$

$$then \;\; e^k p^k \cdot q^k \leq p^k \cdot q^i$$

Then solve for the *minimal* perturbations of the budget hyperplanes, e_*^t, *t=1,2,...,T,* such that the data satisfy $GARP(e_*^t)$. The number, $1-e_*^t$, can be interpreted as the proportion of expenditure that the consumer is allowed to "waste" in situation *t* and still be considered a utility maximizer. Alternatively, the number, e_*^t, measures the efficiency of a consumer in realizing utility-maximizing purchases in situation *t*. This approach can be adapted in a straightforward way to apply to the extended form of the revealed preference axiom, $GARP_{LA}$ (Cox, 1997).

Figure 2 shows results from calculation of Varian's minimal perturbation efficiency index for the consumption good, labor supply, and token balance data (Cox, 1997). There were 263 bundles chosen in the experiment: 243 (or 92.4 %) of the chosen bundles were 100% efficient in maximizing utility; 18 (or 6.8%) of the chosen bundles were 90-99% efficient; and 2 (or 0.8%) of the chosen bundles were 80-89% efficient. Thus 99.2% of the chosen bundles pass a nonparametric "10% test" for consistency with utility maximization.

A conservative test of the utility hypothesis will not count any inconsistency with revealed preference axioms that can be attributed to reporting errors in the data. It is known (Battalio, et al., 1973) that the token economy data is like data from national accounts and consumer panel studies in that it includes reporting errors. The token economy data include both under-reporting and over-reporting of quantities. But the token economy data also include an independent measure of the size of the reporting errors that can be used to adjust tests for the reporting errors. Results from Cox (1997), comparing the partial efficiency measures with the proportional reporting errors for the data are reported in Figure 3. Note that 260 (or 98.9) of the chosen bundles were either fully consistent with utility maximization or contained deviations from full maximization that can be attributed to errors in observations. Only 3 (or 1.1%) of the chosen bundles are known to be inconsistent with utility maximization.

Acknowledgement

Todd Swarthout provided valuable assistance by preparing the table and graphs.

References

Afriat, Sidney. “Efficiency Measures of Production Functions,” *International Economic Review*, 8 (1972): 568-598.

Afriat, Sidney. *Logic of Choice and Economic Theory*. Oxford: Clarendon Press, 1987.

Battalio, Raymond C., *et al*. “A Test of Consumer Demand Theory Using Observations of Individual Consumer Purchases,” *Western Economic Journal*, XI (Dec. 1973): 411- 428.

Becker, Gary S. “Irrational Behavior and Economic Theory,” *Journal of Political Economy*, 70 (February 1962): 1-13.

Cox, James C. “On Testing the Utility Hypothesis,” *Economic Journal*, 107 (July 1997): 1054-1078.

Shafer, Wayne and Hugo Sonnenschein. “Market Demand and Excess Demand Functions.” In *Handbook of Mathematical Economics*, Vol. II, edited by Kenneth J. Arrow and Michael D. Intriligator. New York: North-Holland, 1982.

Sonnenschein, Hugo. “Do Walras' Identity and Continuity Characterize the Class of Community Excess Demand Functions?”, *Journal of Economic Theory*, 6 (August 1973a): 345-354.

Sonnenschein, Hugo. “The Utility Hypothesis and Market Demand Theory,” *Western Economic Journal*, XI (December 1973b) :404-410.

Varian, Hal R. “The Nonparametric Approach to Demand Analysis,” *Econometrica*, 50 (July 1982): 945-973.

Varian, Hal R. “Non-parametric Tests of Consumer Behaviour,” *Review of Economic Studies*, 50 (January 1983): 99-110.

Varian, Hal R. “Revealed Preference With a Subset of Goods,” *Journal of Economic Theory*, 46 (October 1988): 179-185.

Varian, Hal R. "Goodness of Fit for Revealed Preference Tests," unpublished paper, University of Michigan, 1993.

Figure Legends

Figure 1. In order to test the hypothesis that consumers' behavior is consistent with utility maximization, one must use data that are disaggregated (are for individual decision-makers) and complete (include all of the choice variables that enter budget constraints). In order for data from most economies to be complete, it must include observations of consumption goods, labor supplies, and asset net demands. The necessary and sufficient condition for utility maximizing choices is an extended form of the generalized axiom of revealed preference that includes labor supplies and asset net demands. This figure reports the number of observed inconsistencies with this extended form of the axiom (right scale) for 38 consumers in a token economy. Data for 24 consumers contain zero observed violations of the axiom's inequalities. Data for the other 14 subjects contain 54 violations. The figure also reports a measure of the power of the test (left scale) based on a random choice model for which all budget shares (for individual commodities) are equally likely to occur. Zero or low numbers of violations cannot be attributed to low test power for data for those consumers. The number of violations and the test power measure have a -.034 rank correlation.

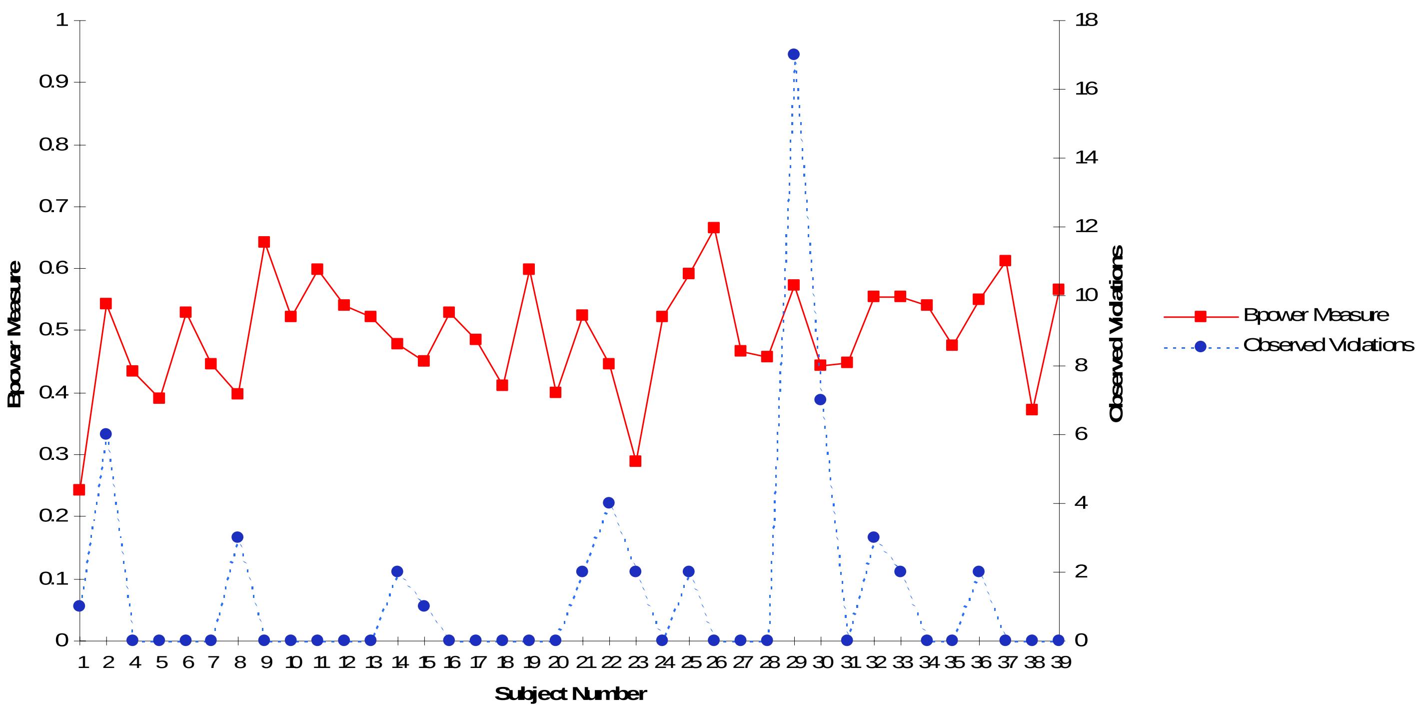

Figure 2. Simply counting the number of inconsistencies with revealed preference inequalities does not inform us about the significance of these observed inconsistencies with utility maximization. A nonparametric measure (partial efficiency) of the significance of the observed inconsistencies is reported in this figure. 243 out of 263 (or 92.4%) of the bundles chosen by consumers in the token economy are 100% efficient in maximizing utility. 18 (or 6.8%) of the chosen bundles are 90 - 99% efficient and 2 (or 0.8%) of the chosen bundles are 80 -89% efficient. 99.2% of the chosen bundles pass a nonparametric "10% test" for consistency with utility maximization.

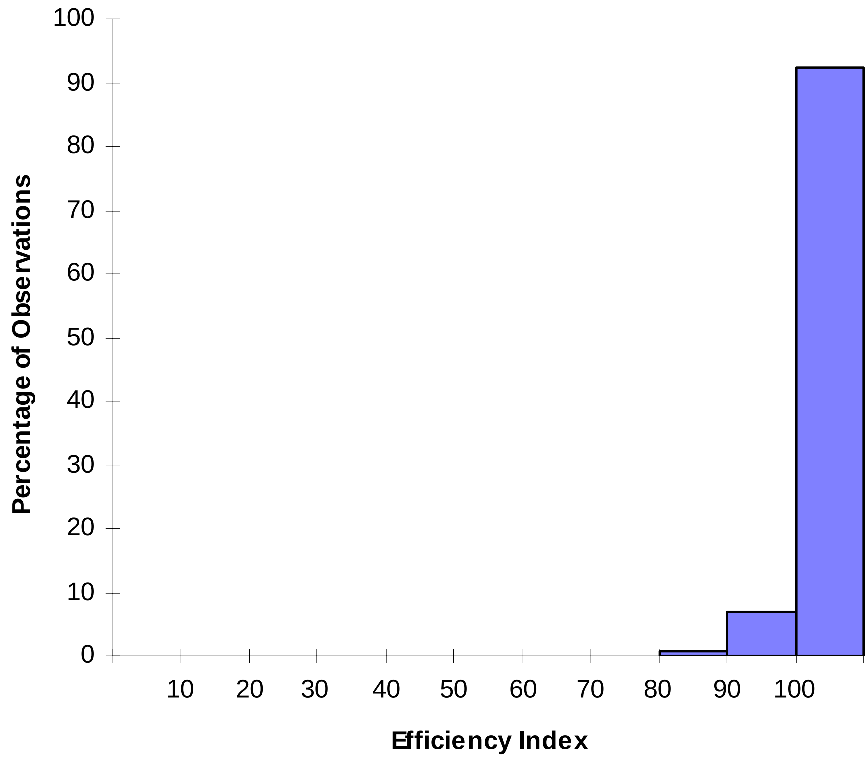

Figure 3. Data for the token economy contain observation errors similar to the observation errors in data for national economies and consumer panel studies. But data for the token economy contain an independent measure of consumer expenditure that makes it possible to discriminate between those observed inconsistencies with utility maximization that can be attributed to observation errors and those that are known to be inconsistencies. This figure reports the results from comparing the partial efficiency measures with the proportional reporting errors for the data. After correcting the tests for these observation errors, only 3 out of 263 (or 1.1 %) of the chosen commodity bundles are known to be inconsistent with utility maximization. 260 out of 263 (or 98.9 %) of the chosen bundles are either fully consistent with utility maximization or contain deviations that can be attributed to errors in observations.

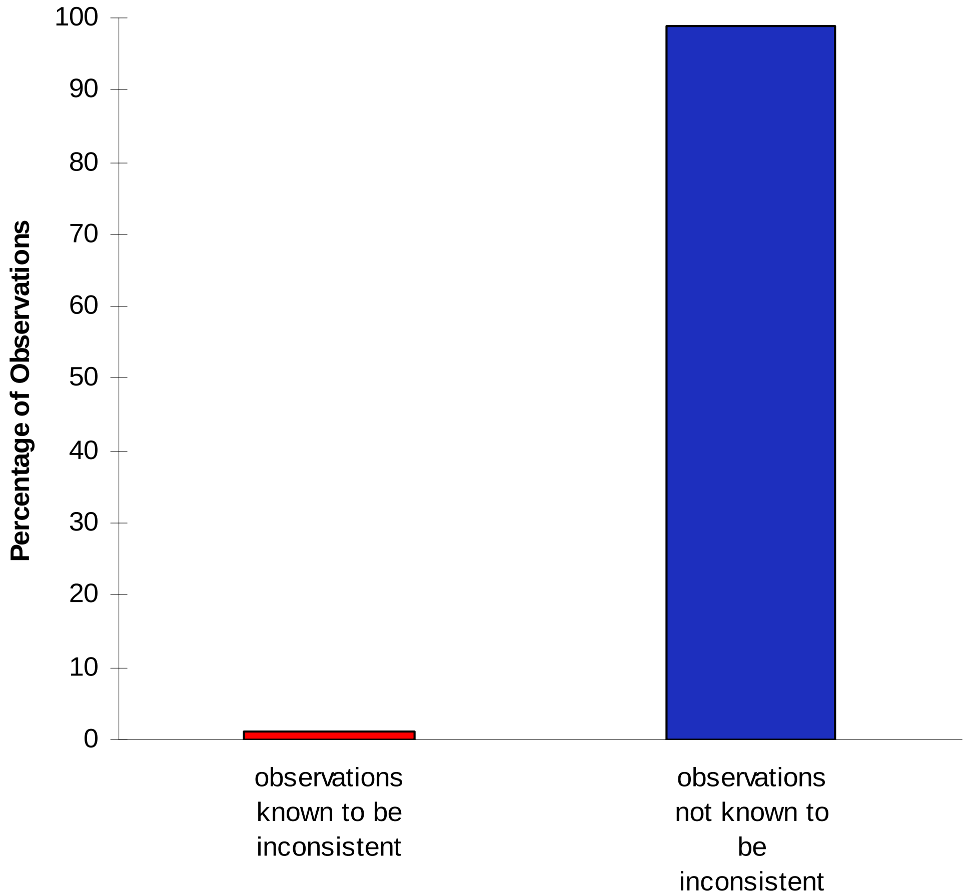

Table 1. Experimental Price Treatments. The experimental treatments were large changes in the prices of Group 1 and Group 2 consumption goods.

Week 1	Group 1	Group 2	Group 3
1	baseline	baseline	baseline
2	0.5 x baseline	2 x baseline	baseline
3	baseline	baseline	baseline
4	2 x baseline	0.5 x baseline	baseline
5	2 x baseline	0.5 x baseline	baseline
6	baseline	baseline	baseline
7	baseline	baseline	baseline