Goodness-of-Fit for Revealed Preference Tests

by

Hal R. Varian
University of Michigan

November, 1990
Current version: July 23, 1993

Abstract. I describe a goodness-of-fit measure for revealed preference tests. This index can be used to measure the degree to which an economic agent violates the model of utility maximization. I calculate the violation indices for a 38 consumers and find that the observed choice behavior is very close to optimizing behavior.

Keywords. Demand analysis, revealed preference

Address. Prof. Hal R. Varian, Department of Economics, University of Michigan, Ann Arbor, MI 48109

Goodness-of-Fit for Revealed Preference Tests

Hal R. Varian

Revealed preference analysis offers necessary and sufficient conditions for choice data to be consistent with the neoclassical model of utility maximization. These conditions, as usually formulated, are exact tests: either the data satisfy the relevant conditions, or they don't.

In many applications it is convenient to have tests for ''almost optimizing behavior.'' Such tests indicate the *degree* to which some data are consistent with an optimization model. In this note I describe two measures for goodness-of-fit of revealed preference conditions based on a construction of Afriat (1987) and describe how to calculate some of these measures. I apply these measures to some consumer demand data collected by Battalio (1973) and show that the subjects in this experiment exhibit behavior that appears to be ''almost optimizing.''

1. Revealed preference notation

Let (p^t, x^t) for $t = 1, \ldots, T$ be a set of prices vectors and consumption bundles. Define the *direct revealed preference* relation R^D by $x^t R^D x$ iff $p^t x^t \geq p^t x$. Let R, the revealed preference relation, be the transitive closure of R^D. The Generalized Axiom of Revealed Preference may be stated as follows:

GARP. *If* $x^t R x^s$ *then* $p^s x^s \leq p^s x^t$.

It is easy to see that if the data (p^t, x^t) were generated by maximizing a nonsatiated utility function, then the data must satisfy $GARP$. Afriat (1967) has shown that if the data satisfy $GARP$, it is possible to construct a well-behaved utility function for which the data are optimizing choices. See Afriat (1967) or Varian (1982) for a proof of these assertions.

If the data violate $GARP$ then there is some pair of observations such that $x^t R x^s$ but $p^s x^s > p^s x^t$. In this case there does not exist a nonsatiated utility function that is consistent with the observed choice behavior.

I would like to thank Eduardo Ley for programming assistance.

2. What does it mean to ‘‘almost’’ satisfy GARP?

However, it is possible that the data is very close to passing GARP; in this case we may not want to reject the maximization hypothesis. In such a circumstance it would be very useful to have an index of the degree of violation of GARP.

The problem is to give precise meaning to the idea that the observed choices are ‘‘close to passing GARP.’’ There have been several attempts to do this in the literature. The first, and in many ways the most satisfactory, was proposed by Afriat (1972). Afriat defines an number called the ‘‘critical cost efficiency index’’ which measures the degree to which a set of data fail to satisfy GARP. We will examine Afriat’s efficiency index below. The main difficulty with this index is that it is a single number for the entire data set. Therefore, it gives little information about which observations are responsible for the revealed preference violations.

Houtman & Maks (1985) suggest finding the largest subset of the data that is consistent with GARP. However, this method discards observations that violate revealed preference, even if they only violate it by a small amount.

Varian (1985) shows how to determine a ‘‘minimal perturbation’’ of the data that is consistent with GARP-like conditions. However, this method is very computation intensive and is impractical for larger data sets.

Gross (1989) suggests partitioning the data sets into a set that satisfies GARP and a set that doesn’t, as in the Houtman and Maks construction. Then he computes a violation index for each observation that doesn’t satisfy GARP using the revealed preference information in the subset of consistent observations. Although Gross’s suggestion is quite interesting, it suffers from the same defect as the Houtman-Maks procedure: observations are deemed not to satisfy revealed preference even if they only violate the conditions by a small amount.

Below I define two new measures of violation that do not have the drawbacks. The indices are analogous to residuals in regression theory in that they are a series of numbers that measure how well each observation ‘‘fits’’ the model---in this case, the model of optimizing behavior. The indices use all of the information in the data, and they are easy to compute.

3. An approximate test of revealed preference

We want to define an approximate version of $GARP$ to describe ‘‘almost optimizing behavior.’’ In order to do so, we follow Afriat (1987). Let e^t for $t = 1, \ldots, T$ be a vector of numbers with $0 \leq e^t \leq 1$. Define $R^D(e^t)$ to be $x^t R^D(e^t) x$ iff $e^t p^t x^t \geq p^s x$. If $x^t R^D(e^t) x$ we say that x^t is directly revealed preferred to x at *efficiency level* e^t. Note that $R^D(1)$ is the standard direct revealed preference relation. Afriat (1987) refers to the numbers (e^t) as the *allowable cost efficiencies.*

Let $R(e^t)$ be the transitive closure of $R^D(e^t)$, and define $GARP(e^t)$ to be

GARP($\mathbf{e^t}$). *If* $x^t R(e^t) x^s$ *then* $e^s p^s x^s \leq p^s x^t$.

If $e^t \equiv 1$ we have the standard GARP test; if $e^t \equiv 0$ we all data trivially satisfy the test. A convenient measure of how close the data are to satisfying GARP is to see how close the e^t can be to 1 and still satisfy $GARP(e)$.

Afriat (1972) proposed using a *uniform* bound. That is, find the largest number e^*, such that $x^t R(e^*) x^s$ implies $e^* p^s x^s \leq p^s x^t$.[1] This number, the *Afriat critical cost efficiency index,* measures how much we have to relax every budget constraint in order for the data to appear to be consistent with maximization. In the interest of brevity, we will call this number the *Afriat efficiency index.*

The construction of this index is illustrated in Figure 1. Here we have a violation of $GARP$: observation 1 is revealed preferred to observation 2 and vice versa. However, the magnitude of the violation is not large: a small perturbation of the budget set through observation 2 removes the violation. In the case depicted the Afriat efficiency index is C/D, a number that is close to 1.

One way to think about the Afriat index e^* is as follows. Think of a direct revealed preference comparison such as that in Figure 1, where $p^1 x^1 \geq p^1 x^2$. Since $p^1 x^2$ is much less than $p^1 x^1$, we are pretty sure that the consumer prefers x^1 to x^2. However, $p^2 x^1$ is only a little bit less than $p^2 x^2$, so we are not so sure that the consumer really prefers x^2 to x^1.

The Afriat index allows the consumer to ‘‘waste’’ a fraction $1 - e^*$ of his income at each observation. If e^* is small, the consumer is wasting very little of his income. If e^* is large, he is wasting quite a lot. In this sense e^* measures the overall ‘‘efficiency’’ of his choice behavior.

[1] This is a slight abuse of notation. By $R(e^*)$ we means the revealed preference relation resulting from setting $e^t = e^*$ for all t.

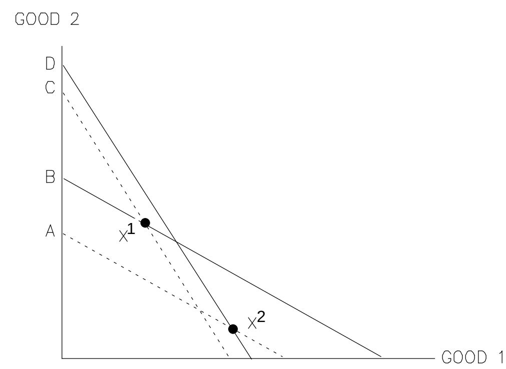

Figure 1. A small violation of revealed preference. A small perturbation of the budget set through observation 2 removes the violation.

The Afriat efficiency index is a nice summary statistic about the overall consistency of the data with optimizing behavior. However, the Afriat index does not give any information about which observations are causing the problems. It would be nice to have a somewhat more disaggregated measure that indicated the minimal amount that one needed to perturb *each* observation in order to satisfy revealed preference conditions.

4. Generalized compensation functions

We approach this problem is a somewhat indirect way. Let S be an *arbitrary* reflexive relation on a set of consumption bundles and define the *generalized compensation function:*

$$m(p, x, S) = \inf_{ySx} py.$$

This function measures the minimal expenditure at prices p to get a bundle that is at least as good as x, according to the relation S.

Let $\succeq$ be a continuous, complete, reflexive, transitive preference ordering of the usual sort. Then $m(p, x, \succeq)$ is essentially the money-metric utility function described in Samuelson (1974) and Varian (1987). Let R be a revealed preference relation that satisfies $GARP$. Then $m(p^t, x^t, R)$ is essentially the overcompensation function in Varian (1982).[2]

[2] Actually, Varian used a somewhat different definition for the overcompensation function called definition the ‘‘approximate overcompensation function,’’ but Knoblauch (1989) has shown that the ‘‘approximation’’ was exact, so that no distinction is necessary.

Now suppose that the relation R is a revealed preference relation that does *not* satisfy $GARP$. As Afriat (1987) points out, the definition of $m(p, x, R)$ still makes sense. Furthermore, since R is reflexive, $m(p, x, R) \leq px$. Suppose that we observe a consumer choosing x^t at prices p^t. If $m(p^t, x^t, R) < p^t x^t$, then the consumer cannot be minimizing his expenditure at observation t. For to say that $m(p^t, x^t, R) < p^t x^t$ is to assert that there exists some x^s such that $x^s R x^t$ and $p^t x^s < p^t x^t$. The *degree* to which the consumer fails to minimize expenditure can be measured by the *violation index*

$$i^t = \frac{m(p^t, x^t, R)}{p^t x^t}$$

If the data are perfectly consistent with $GARP$, we must have $i^t = 1$ for all $t = 1, \ldots, T$. If there is some pair of observations x^t and x^s that violate $GARP$, we will have $x^s R x^t$, but $p^t x^t > p^t x^s$. Hence $i^t < 1$, and the ratio, $i^t = p^t x^s / p^t x^t$ measures the magnitude of this violation.

How does this violation index i^t relate to the idea of ''almost'' satisfying $GARP$ that we discussed earlier? In order to answer this question, it is convenient to establish some properties of $m(p, x, S)$.

Definition. *Let R and S be two arbitrary relations. S is a subrelation of R if xSy implies xRy.*

Fact 1. *If S is a subrelation of R, then $m(p, x, S) \geq m(p, x, R)$.*

Proof. By definition of subrelation, the set $\{y : ySx\}$ is contained in $\{y : yRx\}$. Hence

$$m(p, x, S) = \min_{ySx} py \geq \min_{yRx} py = m(p, x, R).$$

The result now follows. ∎

Fact 2. *If S is a subrelation of R, then the transitive closure of S is a subrelation of the transitive closure of R.*

Proof. If $x^s S x^t$ and $x^t S x^u$ then we must have $x^s R x^t$ and $x^t R x^u$. ∎

Fact 3. *If $f^t \leq e^t$ then $R(f^t)$ is a subrelation of $R(e^t)$.*

Proof. Under the stated condition, if $f^t p^t x^t \geq p^t x$ then $e^t p^t x^t \geq p^t x^t$. Hence if $x^t R^D(f^t) x$ we must have $x^t R^D(e^t) x$. The result now follows from Fact 2. ∎

Proposition 1. *The data satisfy $GARP(i^t)$.*

Proof. Since $R(i^t)$ is a subrelation of $R(1)$, Fact 1 implies

$$i^t = \min_{x^s R(1) x^t} \frac{p^t x^s}{p^t x^t} \leq \min_{x^s R(v^s) x^t} \frac{p^t x^s}{p^t x^t}.$$

Cross multiplying, we see that $i^t p^t x^t \leq p^t x^s$ for $x^s R(i^t) x^t$, which is $GARP(i^t)$ ∎

This says that if we perturb each budget constraint by i^t this will be sufficient to eliminate all revealed preference violations at cost-efficiency level i^t. Hence i^t can serve as a measure of how much the observed choices violation the optimization model.

5. An improved violation index

It is clear from the proof of Proposition 1 that (i^t) is not, in general, not a *minimal* perturbation of the budget sets. If many observations are involved in a cycle the i^t's associated with these observations will all be less than 1. But often perturbing only one of the budget constraints in the cycle will be enough to eliminate the revealed preference violation.

In this section we refine the i^t's described in the last section so as to construct a ''minimal'' perturbation of the budget constraints that will satisfy GARP. Essentially we will start with a vector of 1's, find the revealed preference cycles, and then ''break'' each cycle using the minimum value of i^t over that cycle. We repeat this process until no cycles are left.

The values of i^t that we use will vary as we iterate through the process. Let i^t_k be the value of i^t at the k^{th} iteration. The i^t defined above is i^t_0. Our final output will be a set of numbers (v^t). Again, these will change as we iterate through the algorithm, so we let v^t_k be the value of v^t at stage k.

Given an observation t, and an arbitrary set of numbers (e^t), define $C(t, e^t)$, the *cycle* containing t by

$$C(t, e^t) = \text{all } s \text{ such that } x^s R(e^t) x^t \text{ and } p^s x^s > p^s x^t.$$

If there are no observations that are revealed preferred to x^t and cost less than x^t, then $C(t, e^t)$ is the empty set.

Algorithm for construction (v^t)

0. Initialization stage. Set $k = 0$ and $v^t_0 = 1$ for $t = 1, \ldots, T$.

1. Compute $R(v_k^t)$ and i_k^t. If $GARP(v_k^t)$ is satisfied we are done. Set $v^t = v_k^t$ and terminate.
2. Otherwise, for each observation t find the maximum nonunitary value of i_k^s over all s in the cycle containing t. Let m be the observation where the maximum occurs so that $i_k^m \geq i_k^s$ for all s in $C(t, v_k^t)$ such that $i_k^s < 1$.
3. Let $v_{k+1}^m = i_k^s$. Set $k = k + 1$ and go to step 1.

At each stage the algorithm picks out the largest value of i_k^t that is not 1. Then it uses these maximum of these values over the cycle to ‘‘break’’ each cycle. If $GARP(v_{k+1}^t)$ is not satisfied, it repeats the operations. Eventually all cycles are broken and the algorithm terminates.

Note that breaking all cycles at the first stage doesn't necessarily eliminate all ‘‘subcycles.’’ For example, suppose that $x^1 R x^2 R x^1$ and $x^1 R x^2 R x^3 R x^1$. At the first stage of the cycle it may be that v_0^3 is the largest nonunitary value of the violation index, so that the budget constraint associated with observation 3 is the one that is perturbed. But it may well happen that this perturbation is not sufficient to remove the cycle involving observations 1 and 2, so the algorithm needs to continue to the second stage in order to break this cycle.

The algorithm to calculate (v^t) involves the same sort of calculations necessary to compute (i_0^t) so it is almost as easy to compute. If we want to use a single number to describe the ‘‘efficiency’’ of the choice data, we can use $v^* = \min_t \ v^t$. This gives us an alternative measure of efficiency to the Afriat index described earlier.

6. An example

Here we describe an example of the computation of the indices described earlier. We use the data set collected by Battalio (1973). This consists of observations of 38 long-term patients who operated in a ‘‘token economy’’ at the Central Islip State Hospital. As part of their treatment, these patients worked for tokens which they could redeem for various items such as cigarettes, candy, milk, locker rental, clothes, admission to a dance, etc. During a seven-week period, the relative of various groups of these goods were doubled or halved. Since the prices of some of the goods were halved some weeks and doubled other weeks, prices varied by a factor of four. Data were collected on how the expenditures of each individual responded to the price changes. For more details on the experimental design, see Battalio (1973).

These data have been examined using revealed preference techniques by the original experimenters and by Cox (1989). Battalio *et. al.* used only the observations on purchases of three groups of goods. Cox attempts to take into account labor supply and savings decisions of the group. Both papers find some violations of revealed preference, but the authors argue that the violations are ‘‘small.’’

Battalio (1973) indicate smallness both as a percentage of total expenditure and in terms of measurement error in tokens. For example, they say ‘‘. . . for at least 5 of these subjects, an error of 1 token would have been sufficiently large to prevent [a violation of revealed preference].’’

Cox (1989) used the Afriat efficiency index to measure the degree of violation of revealed preference. He used only 5 weeks of data and found that most of the subjects satisfied $GARP(1)$. The few that failed $GARP(1)$ had high values of the Afriat index, especially compared to his estimate of the measurement error in the data.

In this paper, I apply the notions of ‘‘smallness’’ developed above. Table 1 presents the values of i^t and v_k^t for the subjects who violated $GARP(1)$. The subscript on v_k^t indicates how many iterations it took to converge to a set of v^t’s that satisfied $GARP$. Note that 9 agents out of the 38 exhibited some violations of GARP during the 7-week experiment. The column labeled ‘‘Index’’ give the values of i^t and of v_k^t. The number v_k^t can be taken as an index of the magnitude of the perturbation on expenditure necessary to make the data consistent with $GARP(v_k^t)$. The number under the subject number in the first column is the minimum value of v_k^t over the t’s. As we have seen, this in general is a lower bound on the Afriat number; however, for these data it turned out that $e^* = v^*$ for all subjects.

Note how close to 1 these numbers are. There were 38×7 = 256 bundles chosen in the seven-week experiment. All but 14 of these were fully efficient. Of the inefficient choices, 8 were 97--99 percent efficient, 4 were 93--96 percent efficient, 1 was 91 percent efficient, and one was 81 percent efficient. If we choose a 95% efficiency level as our critical value---perhaps for sentimental reasons---we find that fully 251 of the 256 choices, or 98% percent of the choices were at least 95% efficient.

7. Summary

We have developed two new measures of the goodness-of-fit of revealed preference conditions. One is the violation index i^t, which measures how much each budget constraint has to be perturbed to eliminate to satisfy the revealed preference conditions. The efficiency indices v^t are a refinement of the i^t measures.

We calculated these measures for 38 subjects in a 7-week experiment in which relative prices varied by a factor of 4. We found that this choice behavior was very close to satisfying $GARP$, which is evidence in favor of the neoclassical model of consumer behavior.

Table 1. Calculation of i^t								
Subject	Index	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
1	i^t	0.98	1.00	0.95	0.81	1.00	1.00	1.00
.98	v_2^t	0.98	1.00	1.00	0.98	1.00	1.00	1.00
8	i^t	1.00	1.00	1.00	0.97	1.00	0.70	0.73
.97	v_1^t	1.00	1.00	1.00	0.97	1.00	1.00	1.00
10	i^t	1.00	1.00	1.00	1.00	0.85	0.95	1.00
.94	v_1^t	1.00	1.00	1.00	1.00	1.00	0.94	1.00
12	i^t	0.71	1.00	1.00	0.77	1.00	1.00	0.91
.91	v_2^t	1.00	1.00	1.00	0.97	1.00	1.00	0.91
17	i^t	0.95	1.00	1.00	0.90	1.00	0.81	0.93
.93	v_3^t	0.95	1.00	1.00	0.94	1.00	1.00	0.93
24	i^t	0.95	1.00	1.00	1.00	1.00	1.00	1.00
.99	v_1^t	1.00	0.99	1.00	1.00	1.00	1.00	1.00
28	i^t	1.00	1.00	1.00	1.00	0.92	1.00	1.00
.99	v_1^t	0.99	1.00	1.00	1.00	1.00	1.00	1.00
29	i^t	0.64	0.59	0.94	1.00	0.61	1.00	1.00
.81	v_2^t	1.00	1.00	1.00	1.00	0.81	0.99	1.00
35	i^t	1.00	1.00	1.00	0.75	1.00	1.00	1.00
.99	v_1^t	1.00	1.00	1.00	1.00	1.00	0.99	1.00

References

Afriat, S. (1967). The construction of a utility function from expenditure data. *International Economic Review, 8*, 67--77.

Afriat, S. (1972). Efficiency estimates of production functions. *International Economic Review, 8*, 568--598.

Afriat, S. (1987). *Logic of Choice and Economic Theory*. Oxford: Clarendon Press.

Battalio, R. C. e. a. (1973). A test of consumer demand theory using observations of individual consumer purchases. *Western Economic Journal, XI*, 411--428.

Cox, J. C. (1989). On testing the utility hypothesis. Technical report, University of Arizona.

Gross, J. (1989). Determining the number of violators of the weak axiom. Technical report, University of Wisconsin--Milwaukee.

Houtman, M. & Maks, J. A. (1985). Determining all maximial data subsets consistent with revealed preference. *Kwantitatieve Methoden, 19*, 89--104.

Knoblauch, V. (1989). The money metrics that rationalize a finite set of preference data. Technical report, University of Wisconsin--Milwaukee.

Samuelson, P. A. (1974). Complementarity: an essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. *Journal of Economic Literature, 12*(4), 1255--1289.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica, 50*(4), 945--972.

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

Varian, H. R. (1987). *Microeconomic Analysis.* New York: W. W. Norton & Co.