

Revealed Preference with a Subset of Goods

HAL R. VARIAN*

*Department of Economics, University of Michigan,
Ann Arbor, Michigan 48109*

Received November 4, 1985; revised August 14, 1987

Suppose that you observe n choices of k goods and prices when the consumer is actually choosing from a set of $k + 1$ goods. Then revealed preference theory puts essentially no restrictions on the behavior of the data. This is true even if you also observe the quantity demanded of good $k + 1$, or its price. The proofs of these statements are not difficult. *Journal of Economic Literature* Classification Number: 022. © 1988 Academic Press, Inc.

Suppose that we are given n observations on a consumer's choices of k goods, (p_i, x_i) , where p_i and x_i are nonnegative k -dimensional vectors. Under what conditions can we find a utility function $u: R^k \rightarrow R$ that rationalizes these observations? That is, when can we find a utility function that achieves its constrained maximum at the observed choices?

This is, of course, a classical question of consumer theory. It has been addressed from two distinct viewpoints, the first known as *integrability theory* and the second known as *revealed preference theory*. Integrability theory is appropriate when one is given an entire demand *function* while revealed preference theory is more suited when one is given a finite set of demand observations, the case described above. In the revealed preference case, it is well known that some variant of the Strong Axiom of Revealed Preference (SARP) is a necessary and sufficient condition for the data (p_i, x_i) to be consistent with utility maximization.

Now suppose that we are also given n observations on another chosen good, (z_i) , $i = 1, \dots, n$, where z_i is a nonnegative scalar, but we do not have a price series to accompany these observations. We now ask when will there exist a utility function $u: R^{k+1} \rightarrow R$ that rationalizes the data (p_i, x_i, z_i) for $i = 1, \dots, n$? Equivalently, we can ask when can we find a

* This work was supported by the National Science Foundation. I thank Tom Russell for helpful discussions.

series of scalar prices (q_i) such that the entire data set (p_i, q_i, x_i, z_i) is consistent with utility maximization?

This question is of considerable interest, since we typically can observe only a subset of the goods chosen by a consumer. For example, we would expect that the planned consumption of future goods would enter the utility function, and these are generally not observed. Similarly, contingent consumption plans are also not observed.

The variables (z_i) may also be interpreted as a "demographic" variable such as household size or household location. Such variables are often used in applied demand analysis to control for taste differences. Then we are asking when we can find a family of utility functions, $u(x, z)$, parametrized by z , such that for each fixed z_i the data (p_i, x_i) satisfy the restrictions imposed by demand theory. I will return to this interpretation below.

The first interpretation, that of a missing price, has been addressed in [3] in the intertemporal context, using the machinery of integrability theory. He shows that there are essentially no observable restrictions on demand functions in this context. Here we examine these issues using the methods of revealed preference theory and reach a similar conclusion. However, if one is willing to place bounds on the expenditure on the unobserved good, then we show that demand theory does impose some restrictions on the observed behavior.

1. OBSERVED QUANTITIES, UNOBSERVED PRICES

Let us first describe the form of the revealed preference conditions that we will use. If all prices and goods are observed, a necessary and sufficient condition for these choices to be consistent with utility maximization is that the data satisfy the Generalized Axiom of Revealed Preference (GARP).

DEFINITION. An observation x_i is directly revealed preferred to a bundle x (written $x_i R^0 x$) if $p_i x_i \geq p_i x$. An observation is revealed preferred to a bundle x (written $x_i R x$) if there is some sequence such that $x_i R^0 x_j \cdots x_k R^0 x$. A set of data (p_i, x_i) , $i = 1, \dots, n$, satisfies the Generalized Axiom of Revealed Preference (GARP) if $x_i R x_j$ implies $p_j x_j \leq p_j x_i$.

Further information on these concepts may be found in [1, 2, 5]. GARP is a generalization of the Strong Axiom of Revealed Preference (SARP) that allows for different quantity vectors to be observed for a single price vector. It is therefore appropriate for examining cases where preferences may be weakly convex, rather than strictly convex preferences as is required by SARP.

It is now easy to answer the question posed above. We simply ask when we can construct a price series (q_i) for $i = 1, \dots, n$ such that the entire data set (p_i, q_i, x_i, z_i) satisfies GARP. As it turns out this can be done simply by choosing large enough values for q_i so that the expenditure on the z -good "swamps" the revealed preference comparison. The details are given in the following theorem.

THEOREM 1. *Let (p_i, x_i) , $i = 1, \dots, n$, be a set of data and let (z_i) be a set of n positive scalars. For each i , let E_i be the subset of the data that satisfies $z_t = z_i$ for all t in E_i . Then the only restriction imposed by the maximization hypothesis is that the data (p_i, x_i) satisfy GARP in each subset E_i . In particular, if $z_i \neq z_j$ for all i and j , so that $E_i = \{z_i\}$, then the maximization model imposes no restrictions whatsoever on the observed choices.*

Proof. First we show that the condition is necessary. This follows easily from the fact that $x_s R^0 x_t$ implies that $(x_s, z_i) R^0(x_t, z_i)$ whatever the price q_i . Thus a violation of GARP involving observations in E_i will necessarily create a violation of GARP with the additional good z_i .

In order to prove sufficiency, we will construct a set of prices (q_i) such that the entire data set satisfies GARP. For each i choose q_i such that

$$q_i > \max_{i \notin E_i} \left\{ \frac{p_i(x_j - x_i)}{z_i - z_j}, 1 \right\}. \tag{1}$$

Within each subset E_i we are assured that the data satisfy GARP. What about across subsets? I claim that $(x_i, z_i) R^0(x_j, z_j)$ if and only if $z_i > z_j$, when z_i and z_j are in different subsets E_i and E_j . There are two cases:

(1) $z_i > z_j$. Cross multiplying (1) we have

$$p_i(x_i - x_j) + q_i(z_i - z_j) > 0,$$

which means that $(x_i, z_i) R^0(x_j, z_j)$.

(2) $z_i < z_j$. Cross multiplying (1) again gives

$$p_i(x_i - x_j) + q_i(z_i - z_j) < 0,$$

which means that it is not the case that $(x_i, z_i) R^0(x_j, z_j)$.

Given these choices for q_i , is it possible that the set of data (p_i, q_i, x_i, z_i) could violate GARP? We know that there are no violations within the subsets E_i , so any violations must involve observations from different subsets. But if (x_i, z_i) and (x_j, z_j) are in different subsets, we know that $(x_i, z_i) R(x_j, z_j)$ if and only if $z_i > z_j$ by construction. Thus a violation of GARP would imply $z_i > z_j$ and $z_i < z_j$, which is a contradiction. ■

There are several remarks worth making about this theorem. First, if z_i were a vector, we could simply choose a vector q_i with zeros (or small numbers) in every component but one. The above construction would still work. Second, the numbers z_i provide a complete preference ordering for the subsets E_i . Within each subset, the data are partially ordered by the revealed preference order.

If the variables (z_i) are thought of as demographic variables we can use Afriat's theorem, as described in [5], to construct a piecewise linear utility function, $u(x, z)$, that will rationalize the data in the required sense. Thus, as long as we don't have any violations of revealed preference for fixed values of the demographic variables, the maximization hypothesis puts no restrictions on the behavior of the choice data.

2. OBSERVED PRICES, UNOBSERVED QUANTITIES

The above theorem raises the question of what happens if we observe prices for the omitted good but not quantities? In this case we get no restrictions of any sort.

THEOREM 2. *Let (p_i, x_i) , $i = 1, \dots, n$, be a set of data and let (q_i) , $i = 1, \dots, n$, be a set of positive prices. Then there always exists a set of quantities (z_i) , $i = 1, \dots, n$, such that the data (p_i, q_i, x_i, z_i) satisfy GARP.*

Proof. Choose $z_1 = 0$ and successively define

$$z_{i+1} > \max \left\{ \frac{p_{i+1}x_i - p_{i+1}x_{i+1} + q_{i+1}z_i}{q_{i+1}}, 1 \right\} \quad \text{for } i = 1, \dots, n-1.$$

Then for all $i = 1, \dots, n-1$ we have

$$p_{i+1}x_{i+1} + q_{i+1}z_{i+1} > p_{i+1}x_i + q_{i+1}z_i$$

so that each observation $i+1$ is revealed preferred to observation i . Thus the data must satisfy GARP. ■

Note that the data can be reordered in any way desired so that any preference ordering is consistent with the data. Furthermore, if neither q_i nor z_i is observed there are clearly no restrictions whatsoever on the data (p_i, x_i) .

3. BOUNDING THE EXPENDITURE ON THE OMITTED GOOD

In the constructions given above we have essentially made the expenditure on the omitted good so large that it has "swamped" the revealed

preference comparisons. If we are willing to bound the expenditure on this good, we can get some restrictions on the subset of choices.

THEOREM 3. *Let (p_i, x_i) , $i = 1, \dots, n$, be a set of data and (z_i) , $i = 1, \dots, n$, be an omitted good. Suppose that we postulate that the maximum expenditure possible on the omitted good is bounded by (e_i) , $i = 1, \dots, n$. Then the data (p_i, x_i, e_i) are consistent with utility maximization if and only if there is a positive solution (U_i, λ_i, d_i) to the set of linear inequalities*

$$U_i \leq U_j + \lambda_j p_j(x_i - x_j) + d_j(z_i - z_j)$$

$$d_j z_j \leq \lambda_j e_j.$$

Proof. Choose $q_i = d_i/\lambda_i$ and rewrite the inequalities in the form

$$U_i \leq U_j + \lambda_j p_j(x_i - x_j) + \lambda_j q_j(z_i - z_j)$$

$$q_j z_j \leq e_j.$$

But these are simply the Afriat inequalities which have been shown to be a necessary and sufficient condition for utility maximization in [1, 2, 5]. ■

In order for this result to be of interest, we have to show that the inequality conditions given in Theorem 3 are not vacuous. A simple example will suffice.

Suppose, for example, that we have two observations on three goods with the following specifications: $x_1 = (1, 2)$, $x_2 = (2, 1)$, $p_1 = (1, 2)$, $p_2 = (2, 1)$, $z_1 = 1$, and $z_2 = 2$. It is easy to see that these observations violate the Generalized Axiom of Revealed Preference. Suppose that we try to patch things up by choosing prices (q_1, q_2) so that observation 2 is not revealed preferred to observation 1. This implies that

$$5 + 2q_2 \leq 4 + q_2,$$

which yields

$$q_2 \leq -1,$$

which is impossible.

Since that doesn't work, let's try to assure that observation 1 is not revealed preferred to observation 2. In this case we have

$$5 + q_1 \leq 4 + 2q_1,$$

which implies that $q_1 \geq 1$. Thus the expenditure on z_1 must be at least 1 in order to satisfy the revealed preference restrictions. It follows that if the

bound on the expenditure on the omitted good is less than 1, the data cannot be consistent with GARP.

If (q_i) is observed but (z_i) is not observed, a similar set of inequalities can be constructed, but they are now nonlinear. This is also the case if total expenditure is observed, but neither q_i nor z_i is observed.

Theorem 3 suggests a way to check for "significant violations" of revealed preference. Suppose that we have a set of data (p_i, x_i, z_i) that violates revealed preference. It would be convenient to have an "index" of the degree of violation of revealed preference. One way to do this would be to find the smallest value \bar{e} such that the inequalities

$$U_i \leq U_j + \lambda_j p_j(x_i - x_j) + d_j(z_i - z_j)$$

$$d_j z_j \leq \lambda_j \bar{e}$$

have a positive solution. The number \bar{e} tells us how much the expenditure would have to be on the z -good in order to satisfy the restrictions of utility maximization. If the z -good is a demographic variable, \bar{e} would give us an index of how important it would have to be for it to account for the taste differences necessary to describe the data. Since the inequalities described in Theorem 3 are linear in the unknown variables, checking for feasibility does not pose undue computational difficulties.

4. CONCLUSION

If the utility function is assumed to have some special structure such as separability—where $u(x, z)$ has the form $U(v(x), z)$ —it is well known that maximization does impose restrictions on the data (p_i, x_i) . These restrictions are summarized in [6]. However, without this assumption of special structure, there are essentially no restrictions imposed by the maximization model on a subset of the choice data.

I take this to be a negative result, similar in spirit to the Sonnenschein–Mantel–Debreu results described in [4], although obviously not as deep. The sad fact of the matter is that the restrictions imposed by the optimization hypothesis only apply when we have observed the entire choice set. Hence the normal sorts of tests for consistency of observed choice must be interpreted instead as tests for separability of the observed choices from other variables in the utility function rather than test of maximization *per se*.

REFERENCES

1. S. AFRIAT, The construction of a utility function from expenditure data, *Int. Econ. Rev.* **8** (1967), 67–77.
2. E. DIEWERT, Afriat and revealed preference theory, *Rev. Econ. Stud.* **50** (1973), 419–426.
3. H. POLEMARCHAKIS, Expectations, demand, and observability, *Econometrica* **51** (1983), 565–574.
4. W. SHAEFER AND H. SONNENSCHNEIN, Excess demand functions, in “Handbook of Mathematical Economics” (K. Arrow and M. Intrilligator, Eds.), Vol. 2, North-Holland, Amsterdam, 1983.
5. H. VARIAN, The nonparametric approach to demand analysis, *Econometrica* **50** (1982), 945–973.
6. H. VARIAN, Nonparametric tests of consumer behavior, *Rev. Econ. Stud.* **50** (1983), 99–110.