Staff Papers Series

Staff Paper P89-14 March 1989

ON TESTING FOR
REVEALED PREFERENCE CONDITIONS

March 1989

by

Yacov Tsur
University of Minnesota

Department of Agricultural and Applied Economics

University of Minnesota
Institute of Agriculture, Forestry and Home Economics
St. Paul, Minnesota 55108

On Testing for Revealed Preference Conditions

Yacov Tsur*

Department of Agricultural and Applied Economics, University of Minnesota

Abstract

A procedure to test for the significance of violations of revealed preference conditions is described. The procedure is simple and hence may especially be appropriate for large data sets. An application to consumption data is presented.

*Department of Agricultural and Applied Economics, University of Minnesota, 1994 Buford Avenue, St. Paul, MN 55108.

The helpful comments of Ted Graham-Tomasi are gratefully acknowledged

Introduction

Applications of nonparametric revealed preference theory, while free of misspecification errors, are generally not free of measurement errors. Data may be viewed as being generated from an underlying true, unobserved structure and some measurement error process. Suppose the data violate GARP (Varian, 1982, p. 947). A question then arises as to whether this is a significant indication that the true underlying structure violates GARP or is it likely that the true structure satisfy GARP and the violations in the observed data were caused by the measurement errors?

This note describes a procedure designed to test for the significance of GARP violation. Its implementation is simple, requiring only a few, fast computations. Thus, it may in particular be useful when dealing with large data sets, possibly as a screening test before a more elaborate method, such as that suggested by Varian (1985), is considered. The present approach provides a natural interpretation of Afriat's (1967) efficiency indexes, as defined in Varian (1987), in terms of the measurement errors.

Specification of the Data Generating Process

For k goods and n time periods let X^i and P^i be the k by 1 vectors of observed quantities and prices, respectively, at time period $i=1,2,..n$. The observations contain errors; the corresponding true, unobserved, quantities and prices are denoted by X^{*i} and P^{*i}. Let $C(i,j) = P^i \cdot X^j$ be the observed expenditure of goods consumed at time j in terms of period i prices. In a similar manner define $C^*(i,j) = P^{*i} \cdot X^{*j}$. The revealed preferred relation is indicated by R (see Varian 1982, p. 947). The equivalent concept for the true structure is indicated by R^*; thus $X^{*i}R^*X^{*j}$ is interpreted as "$X^{*i}$is preferred to X^{*j}" according to the unobserved quantities and prices (the term "revealed" is dropped since X^{*j} and P^{*1} are unobserved). The starred variables P^{*j}, X^{*j} and C^* will be referred to as the true structure.

The data $\{X^j, P^j,\ j=1,2,..,n\}$ satisfy GARP if X^iRX^j implies $C(j,j) \le C(j,i)$ for all i,j. The true quantities and prices satisfy GARP if $X^{*i}R^*X^{*j}$ implies $C^*(j,j) \le C^*(j,i)$ for all i,j. Varian (1987) defines another revealed preference condition: the data $\{X^j, P^j, e^j,\ j=1,2,..,n\}$ satisfy GARPe if $X^iR_eX^j$ implies $e^jP^j \cdot X^j \le P^j \cdot X^i$; where R_e^o is defined as $X^jR_e^oX$ iff $e^jP^jX^j \ge P^jX$, R_e is the transitive closure of R_e^o and the e^j are n scalars satisfying $0 \le e^j \le 1$. The set $\{X^j, P^j, e^j,\ j=1,2,..,n\}$ will be called the perturbed data set and $e = (e^1, e^2, ..., e^n)$ is the perturbation vector.

It is assumed that the observed quantities and prices are related to the true structure according to $X^j = V_jX^{*j}$ and $P^j = W_jP^{*j}$, where V_j and W_j are scalar random variables such that $v_j = \log V_j$ and $w_j = \log W_j$ are independently and identically distributed (iid) normal variates with zero means and variances given by σ_v^2 and σ_w^2, respectively. Letting $c = \log C$ and $c^* = \log C^*$, it follows that

$$c(i,j) = c^*(i,j) + v_i + w_j; \quad v_i + w_j \sim_{iid} N(0, \sigma_v^2 + \sigma_w^2),$$

which characterizes the data generating process. In particular, for i=j,

$$c(j,j) = c^*(j,j) + \epsilon_j; \quad \epsilon_j \sim_{iid} N(0, \sigma^2),\ j=1,2,...,n, \tag{1}$$

where $\epsilon_j = v_j + w_j$ and $\sigma^2 = \sigma_v^2 + \sigma_w^2$. The n by 1 vector which consists of the diagonal elements of $c(\cdot,\cdot)$ [resp. $C(\cdot,\cdot)$] will be indicated by c [resp. C] and will be refer to as the expenditure vector.

The Test Procedure

Our test concerns observations which violate GARP. We make use of the idea set forth by Varian (1985) and perturb the data until GARPe is satisfied. We then inquire whether it is plausible that the perturbed data represent a feasible true structure from which the actual data were generated. Following Afriat (1967) and Varian (1987), perturbations are allowed only on the diagonal elements of the expenditure matrix, i.e. on the expenditure vector C.

Let the distance between any two vectors in R^n be defined as the square of the Euclidean norm of the difference vector divided by n. For $\rho \geq 0$, let $T_n(\rho) = \{u \in R^n: \Sigma_j [u_j - c_j)]^2/n \leq \rho\}$ be the set of all n by 1 vectors that are at most distance ρ away from the expenditure vector c. The set $T_n(\rho)$ is such that $T_n(\rho') \subseteq T_n(\rho")$ whenever $\rho' \leq \rho"$, with $T_n(0) = \{c\}$ and $T_n(\infty) = R^n$.

Any $u \in R^n$ has a perturbation vector attached to it defined by $e^j = \exp(u_j - c_j)$ such that $U_j = \exp(u_j)$ is derived from the expenditure vector C according to $U_j = e^j C_j$, j=1,2,..,n. In this way one can attach a perturbed data set to any vector $u \in R^n$.

Definition: *$T_n(\rho)$ satisfies GARPe if there exists $u \in T_n(\rho)$ such that the perturbed data generated by $e^j = \exp[u_j - c_j]$, j=1,2,..,n, satisfies GARPe.*

Clearly $T_n(0)$ does not satisfy GARPe (since it entails $e^j = 1$ and GARP is violated by the data) whereas $T_n(\infty)$ vacuously satisfies GARPe (setting $u_j = -\infty$ implies $e^j = 0$). Moreover, if $T_n(\rho')$ satisfies (does not satisfy) GARPe then the same holds true for any $T_n(\rho)$ with $\rho \geq (\leq) \rho'$.

The true expenditure vector c^* is of distance [cf. equation (1)]

$$s_n^2 = \sum_{j=1}^{n} \epsilon_j^2/n$$

away from c. Define $\rho_n = \text{Min}\{\rho: T_n(\rho) \text{ satisfies GARPe}\}$ and suppose $\rho_n \leq s_n^2$. Then it must be that $T_n(s_n^2)$ satisfies GARPe as well. Obviously $T_n(s_n^2)$ contains c^*, which means that the possibility that the true expenditure vector satisfies GARP cannot be ruled out. Under such circumstances we shall say that the violation of GARP by the observed data is not sufficiently severe to imply that the true structure violates GARP as well. On the other hand, if $\rho_n > s_n^2$, then any set $T_n(\cdot)$ that satisfies GARPe cannot possibly contain c^* and we shall interpret this as evidence that the true structure violates GARP.

In view of the above, the null hypothesis, maintaining that the true structure satisfies GARP, is specified as:

$$H_o: \ \rho_n \le s_n^2.$$

Consider the simple hypothesis

$$H_{oo}: \ \rho_n = s_n^2$$

and note that H_{oo} is contained in H_o and rejecting H_{oo} in favor of the alternative $\rho_n > s_n^2$ implies rejecting any other simple hypothesis contained in H_o. Now $s_n^2 = \sum_{j=1}^{n} \epsilon_j^2/n \sim (\sigma^2/n)\chi_n^2$ [cf. equation (1)]. Hence, under H_{oo},

$$n\rho_n \sim \sigma^2\chi_n^2.$$

The test is performed, utilizing aspects of Varian's (1985) procedure, in the following fashion: i) calculate $\hat{\rho}_n \ge \rho_n$ [see procedure below]; ii) calculate $\bar{\sigma}^2 = n\hat{\rho}_n/\chi_g^2(\alpha)$, where $\chi_n^2(\alpha)$ is defined from $\Pr\{\chi_n^2 \ge \chi_n^2(\alpha)\} = \alpha$; iii) reject H_o, at α percent significance level, if it is believed that $\sigma^2 < \bar{\sigma}^2$. When σ^2 is known a priori, H_o is rejected if $n\hat{\rho}_n/\sigma^2 > \chi_n^2(\alpha)$.

It remains to calculate $\hat{\rho}_n$, which is done by the following algorithm.

Input: the n by n expenditure matrix $C(\cdot,\cdot)$

Output: a perturbed data set satisfying GARP$_e$ and a distance index $\hat{\rho}_n \ge \rho_n$.

1) set $M(j)=1$ and $C_e(i,j)=C(i,j)$ for $i,j=1,2,..,n$;

2) set $C_e(j,j)=M(j)C(j,j)$, $j=1,2,..,n$;

3) for $i,j=1,2,..,n$ set $R_e^o(i,j)=1$ or 0 as $C_e(i,i) \ge$ or $< C_e(j,i)$, respectively, and calculate its transitive closure R_e [for an algorithm to calculate the transitive closure of a matrix see Varian (1982, p. 972)];

4) set $G_e=\{j: R_e(i,j)=1$ and $C_e(j,j) > C_e(j,i)$ for at least one case $i\}$;

5) if $G_e = \emptyset$ then go to 7, else go to 6;

6) calculate the n by 1 vector M as

$$M_j = \begin{cases} \min_{x^i R_e x^j} \{C(j,i)/C(j,j)\} & ; j \in G_e \\ M_j & ; j \in G_e^c \end{cases}$$

(G_e^c indicating the complement of G_e) and go to 2;

7) calculate

$$\hat{\rho}_n = \sum_{j=1}^{n} [\log M_j]^2/n \qquad (2)$$

and stop.

This procedure can take at most n iterations, since once any case j has been corrected it can never violate GARP$_e$ again. Clearly, the perturbed data generated by the perturbation vector e = M satisfy GARP$_e$. The j'th element of the associated perturbed expenditure vector is $C_e(j,j)=e^j C(j,j)$. It is easy to verify that the distance of the log of this vector from the log of the actual expenditure vector, c, is given by $\hat{\rho}_n$ of equation (2). Thus $\rho_n \le \hat{\rho}_n$. (It may be possible to design a procedure that calculates ρ_n itself; in the spirit of simplicity, however, we shall not pursue this task here.) With $\hat{\rho}_n \ge \rho_n$, the test is more conservative in the sense that if H_0 cannot be rejected with $\hat{\rho}_n$ it obviously will not be rejected with ρ_n.

Application

Our data set contains consumption and price data of four (k=4) major meat types in Spain for the 150 months (n=150) of the period January 1970 through July 1982. The mean and variance of the expenditure sample (c_j = log C(j,j), j=1,2,..,n) are 16.634 and 0.0243, respectively. A computer code realization (fortran) of a GARP test, based on the algorithms described in Varian (1982), detected 33 cases (months) in which at least one GARP violation occurs. A perturbation vector e = M that satisfies GARP$_e$ was constructed according to the procedure described in the previous section (it took one iteration to satisfy GARP$_e$) and provided the distance index $\hat{\rho}_n = \sum_{j=1}^{n}[\log M_j]^2/n = .000046$.

Choosing $\alpha = .01$ and $.1$ yields $\chi^2_{150}(.01) = 192.431$ and $\chi^2_{150}(.1) = 172.482$ which imply $\bar{\sigma}^2 = n \cdot \hat{\rho}_n / \chi^2_n(\alpha) = 7.9\times10^{-6}$ and 8.8×10^{-6}, respectively.

As noted above, the variance of the sample $\{c_j,\ j=1,2,..,n\}$ is 0.0243. Even if c^*_j, $j=1,2,..,n$, explain 99 percent of this variance we are still left with an error variance which is of order of magnitude of 10^{-4}. It therefore appears unreasonable to suppose that the variance of the (log of) actual expenditures is less than $\bar{\sigma}^2$. Thus H_0 cannot be rejected. We concluded that the violation of GARP by the data is not sufficiently large to imply that the true structure violate GARP as well.

References

Afriat, Sydney N., 1967, The construction of utility function from expenditure data, *International Economic Review* 8, 67-77.

Varian, Hal R., 1982, The nonparametric approach to demand analysis, *Econometrica* 50, 945-972.

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

Varian, Hal R., 1987, On goodness-of-fit of revealed preference conditions, CREST Working Paper, Department of Economics, University of Michigan.