A Household-Based, Nonparametric Test of Demand Theory

Melissa Famulari

The Review of Economics and Statistics, Vol. 77, No. 2 (May, 1995), 372-382.

Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28199505%2977%3A2%3C372%3AAHNTOD%3E2.0.CO%3B2-R

A HOUSEHOLD-BASED, NONPARAMETRIC TEST OF DEMAND THEORY

Melissa Famulari*

Abstract—A nonparametric test of demand theory, determining the consistency of price-expenditure data with the Generalized Axiom of Revealed Preference (GARP), is applied to household microeconomic data for the first time. The fraction of consumption bundle comparisons that violate GARP is proposed as a summary statistic to indicate the consistency of a data set with revealed preference axioms. This violation rate is calculated among consumption pairs with similar expenditures to increase the power of the test. The results suggest that similar households' consumption choices and budget constraints are generally consistent with the joint hypotheses of optimizing behavior and common preferences.

I. Introduction

THERE are numerous empirical studies, both parametric and nonparametric, that test the consistency of consumption choices with demand theory.[1] Parametric tests have generally rejected the restrictions of demand theory: a symmetric and negative semidefinite Slutsky matrix and homogeneity.[2] However, since parametric studies jointly test the hypothesized functional form of preferences and utility maximization, it is not possible to distinguish rejection of demand theory from rejection of the functional form.[3] Nonparametric methods directly test for optimizing behavior by examining price–consumption data for violations of the Generalized Axiom of Revealed Preference (GARP). Nonparametric tests for the consistency of commodity choices with demand theory have been applied to aggregate time series data and have found few inconsistent choices.[4] Results using aggregate data are less compelling because all idiosyncratic aspects of choice have been averaged out.[5]

In this study I test for optimizing behavior in commodity choices using repeated cross sections of households from the 1982–1985 Consumer Expenditure Survey (CE). The data combine household demographic characteristics and annual expenditures on eight aggregate categories of goods and a composite "other" category with interarea–intertemporal price indices for 25 major metropolitan areas. These data make it possible to test the consistency of household consumption with revealed preference axioms. Moreover, the CE reports actual consumption choices for individual decision-making units. Sample sizes are larger and the variation in budget shares is greater than in aggregate consumption data sets. As a result, these tests of GARP across individual households with similar total expenditures are stronger tests of demand theory than previous tests using aggregate data.

This paper also addresses two methodological issues that arise in applying nonparametric methods: the exactness of the test of GARP and measurement error. The test of GARP is exact: one inconsistent choice leads to the rejection of optimizing behavior. In practice, some violations of GARP are likely to be found in most data sets. Therefore, determining whether inconsistent choices are significant violations of the theory has been an important topic of research (Banker and Maindiratta (1988), Gross (1990), Varian (1990)). Previous research has generally focused on either the number of pairs of consumption bundles that violate demand theory or the number of observations involved in violations.

In a sample of N consumption choices there are $N*(N-1)$ consumption bundle comparisons that could conceivably violate GARP. If the vast majority of comparisons fail to reject GARP, then proceeding to estimate a parametric, static

Received for publication May 21, 1992. Revision accepted for publication May 16, 1994.

* University of Texas at Austin.

The author conducted this research, in part, at the Bureau of Labor Statistics. I would like to thank Stephen Bronars for numerous insightful comments. Marilyn Manser, Marshall Reinsdorf, John Ruser, and two anonymous referees made helpful suggestions as well. Leslie Platt provided excellent research assistance.

[1] These studies, as well as the current study, assume that consumption is separable from leisure.

[2] See the studies cited in Deaton (1986) and Blundell et al. (1993).

[3] However, Deaton (1986) argues that functional form is not likely to be the source of the failure of demand theory because there have been many studies using many different functional form assumptions.

[4] See Landsburg (1981), Swofford and Whitney (1987), Manser and McDonald (1988), and Varian (1990).

[5] Aggregation may cause parametric studies to reject demand theory while making average consumption choices less likely to reject GARP by averaging out individual choices. In fact, Becker (1962) showed that, on average, random behavior will be consistent with revealed preference axioms.

utility model seems warranted. However, there are some comparisons where GARP is unlikely to be rejected, regardless of how consumption choices are made. If total expenditures (in constant dollars) are widely different across two periods, the consumption choice in the high expenditure period will generally be preferred over the low expenditure period even if preference orderings are quite different across the two periods. Therefore, the fraction of consumption bundle comparisons with similar total expenditures (in constant dollars) that violate GARP is proposed as a measure of the consistency of a data set with revealed preference theory.[6]

Researchers have recognized that measurement error and small optimization errors may lead to violations of GARP, and have proposed several procedures for determining whether or not these types of errors could account for observed outcomes of nonparametric tests. Afriat (1967) proposed that consumption bundles that have similar total cost at given prices were too alike for individuals to determine which bundle is preferred and so their direct comparison should provide no information about preferences. However, as noted by Gross (1990) and Varian (1990), even though at the same set of prices two consumption bundles have approximately the same cost, the underlying consumption choices may be quite different. More importantly, removing the preference information obtained from consumption bundle pairs with similar total expenditures reduces the power of the nonparametric test of demand theory. In this paper, an individual is assumed to be unable to rank order two consumption bundles if: (i) both consumption bundles have similar total cost at constant prices and (ii) the budget share of each good at constant prices is similar across consumption bundles. This modification avoids the problems inherent in Afriat's original proposal, but maintains the advantages of Afriat's approach; computational costs are low and an inappropriate assignment of preference will not "cause" a series of violations of GARP.

The key limitation of any cross-section empirical model of individual choice is that we must assume preferences are homogeneous across households. In this paper, some heterogeneity in preferences is allowed by grouping households into 43 different subsamples based on their demographic characteristics. Each subsample was tested separately for consistency with GARP. One demographic group was clearly an outlier and rejected GARP. Across the remaining 42 separate tests of GARP, the median violation rate is 3.87% among consumption bundle comparisons where the difference in total expenditures (in constant dollars) is less than 10%.

The nonparametric tests suggest that similar households' commodity choices are generally consistent with the joint hypotheses of common preferences and optimizing behavior. I find that allowing for differences in preferences across demographic groups improves the consistency of these data with revealed preference axioms. However, all the violations of revealed preference that do occur cannot be attributed to measurement error. Despite this caveat, there exists a utility function that could nearly rationalize these data and thus estimating a parametric demand system with these data is reasonable.

II. Nonparametric Test for Preference Maximization

Assume we observe a price vector $P_i = \{P_{1i}, P_{2i}, \ldots, P_{mi}\}$ and a quantity vector $X_i = \{X_{1i}, X_{2i}, \ldots, X_{mi}\}$ for time periods (or individuals) $i = 1$ through N. Expenditures in period i can be expressed as $P_i X_i = \sum_k P_{ki} X_{ki}$, for goods $k = 1, \ldots, m$. Arbitrarily choose P_1 to be the base price vector. Then $M_i = \sum_k P_{k1} X_{ki}$ equals total expenditures in base period dollars. Define the direct revealed preference operator (*DRP*) as follows:

$$i\ DRP\ j \text{ if and only if } P_i X_i \geq P_i X_j. \qquad (1)$$

In addition to a *direct* revealed preference relationship, we can define a revealed preference (*RP*) operator. Let $i\ RP\ j$ if and only if there exists a sequence of periods $s \ldots z$ such $i\ DRP\ s$, $s\ DRP\ r, \ldots, z\ DRP\ j$. The Generalized Axiom of Revealed Preference (GARP) can be stated as $i\ RP\ j$ implies $P_j X_j \leq P_j X_i$. Varian (1982) proves that if a set of data is consistent with GARP, then that data can be "rationalized" by a nonsatiated, continuous, concave, monotonic utility function and outlines an algorithm for testing GARP.

[6] Whenever total expenditures are compared in the paper, the comparison is in *constant dollars*.

III. Violations of GARP

A. *The Violation Rate*

There is no error term in Varian's nonparametric tests of GARP. Therefore, strictly speaking, if even one pair of observations violates GARP, preference maximization is rejected. Given measurement error, it seems unreasonable to specify the acceptance region of the test as 100% consistency with GARP. A violation of GARP can occur either because of non-optimizing behavior or because of measurement error. In addition, increasing the sample size does not affect the rejection criterion even though the potential number of times GARP could be violated rises approximately exponentially with the number of observations.

To determine whether a data set is consistent with the hypothesis of utility maximization, it seems useful to compare the number of times the data fail to reject GARP to the number of GARP violations. If consumption choices and budget constraints are consistent with GARP the vast majority of the time, then proceeding to fit the data with a common, static utility function seems warranted. Alternatively, if consumption choices frequently violate GARP, we should question inferences drawn from a parametric model, even if the estimated parameters appear to be consistent with utility maximizing behavior. For this reason tests of GARP are an important precursor to parametric estimation.

If there are N observations, then there are $N*(N-1)$ pairwise comparisons of consumption bundles that either are consistent with or violate GARP. Define an $N \times N$ matrix, A, and let $[a_{ij}] = 0$ when i not *RP* j and j not *RP* i, $[a_{ij}] = 1$ if i *RP* j and j *DRP* i (i and j violate GARP), and $[a_{ij}] = 2$ otherwise. The number of violations, V, equals the number of terms in the A matrix where $[a_{ij}] = 1$. The number of GARP consistent comparisons, C, is defined to be the number of terms in the A matrix where $[a_{ij}] = 2$ and $i \neq j$. Define the violation rate to be $V/(V+C)$.[7] The violation rate is a summary statistic that indicates the consistency of a set of data with the hypothesis of optimizing behavior.[8]

Although counting the pairs of consumption bundles that violate GARP is straightforward using Varian's (1982) algorithm, determining the number of violations of utility maximization can be problematic. Consumption bundle comparisons that violate revealed preference axioms are not independent. For example, suppose the *DRP* relationships among four consumption bundles are as follows: i *DRP* j, j *DRP* k, k *DRP* l, and l *DRP* i. Since there is a cycle of preference leading to i *RP* l, the comparison of i and l violates GARP. This same cycle of preference also leads to j *RP* i and so the comparison of j and i violates GARP. In fact, anywhere one starts on the *RP* cycle leads to a pair of consumption bundles that violate GARP, i.e., $[a_{il}] = 1$, $[a_{ji}] = 1$, $[a_{kj}] = 1$, and $[a_{lk}] = 1$. Although there are four pairs of bundles that violate GARP, in what sense are there four separate violations of utility maximization? Moreover, the comparisons of l with i, i with j, j with k, and k with l do not violate revealed preference. In what sense are these comparisons consistent with utility maximization? An alternative measure for the violation rate would be to calculate the fraction of revealed preference cycles in a data set that lead to GARP violations. In the example above, there are five cycles of preference in total (four *DRP* cycles and one *RP* cycle) and one (the *RP* cycle) leads to a violation of GARP. However, the computational costs of this method of counting violations increase approximately exponentially with sample sizes, and would be prohibitive in data sets with more than 30 observations. For this reason I focus on consumption bundle pairs in calculating the GARP violation rate.[9] The advantages of this approach are that (i) it is feasible to compute in large data sets and (ii) it indicates the overall consistency of a set of data with the hypothesis of utility maximization.

[7] Pairwise comparisons where $[a_{ij}] = 0$ are not included in the violation rate because the number of such comparisons will increase when the possibility of measurement error is introduced into tests of GARP below. In practice, only 1% of the total number of pairwise comparisons had $[a_{ij}] = 0$.

[8] The properties of the test statistic proposed here are not known both because the source and distribution of the randomness in the model has not been specified. Furthermore, pairwise comparisons of consumption bundles are not independent, so central limit theorems cannot be invoked.

[9] In the four observation example provided in the text, the GARP violation rate is 50%: there are twelve pairwise comparisons of commodity bundles that could violate GARP, four pairs violate GARP, four pairs are consistent with GARP, and four pairs where neither consumption bundle is revealed preferred.

B. Increasing the Power of the Test

The number of GARP consistent comparisons may be misleading because it is possible for two consumption bundles to be trivially consistent with GARP. Suppose an individual would need only one-fifth of her current income to afford a consumption bundle she previously chose. Whether or not the individual's preferences changed, she is likely to prefer her current consumption bundle to the bundle that costs one-fifth as much. As a result, this comparison is unlikely to lead to a violation of GARP, and so adds little to the power of the test. To increase the power of tests of GARP, I calculate the violation rate separately among pairs of consumption bundles where the difference in constant price expenditures is small. Define consumption bundles i and j to have similar total expenditures if:

Definition 1: i and j have similar total expenditures if and only if

$$2*(M_i - M_j)/(M_i + M_j) \leq K$$

where the researcher chooses K. Define an $N \times N$ matrix, Z, and let $[z_{ij}] = [a_{ij}]$ when i and j have similar total expenditures and $[z_{ij}] = 0$ otherwise. Again the violation rate is calculated as $V/(V + C)$ where V and C are now defined by elements in the Z matrix.

There are now two parameters, chosen by the researcher, which determine the rejection region for the violation rate: (i) the critical value of the violation rate and (ii) the value of K that determines the number of comparisons with similar total expenditures. A violation rate of 5%, a standard significance test critical value, is probably a reasonable criterion for the rejection of GARP. A reasonable value for K, the cutoff value for expenditure differences, is less clear. Higher values of K result in less powerful tests of GARP; we are more likely to accept the hypothesis of utility maximization regardless of the true model of consumer choice. Low values of K increase the likelihood of incorrectly rejecting the null hypothesis due to measurement error. Since the tradeoffs between type I and type II errors are uncertain, the best approach may be to report the violation rate for a range of values of K.[10] However, if consumption bundle comparisons with expenditure differences as large as 20% violate GARP more than 5% of the time, then it seems reasonable to reject the hypothesis of preference maximization in these data.

C. Afriat's Efficiency Index, Expenditure Reallocation, and Measurement Error

Measurement error can confound the nonparametric analysis when only a slight difference in one of the measured quantities or prices in a pair of consumption bundles would change the preference ordering obtained from their direct comparison. Therefore, in the presence of measurement error, the researcher may have less confidence in the *DRP* relationships established by the comparison of similar consumption bundles. The direct revealed preference relationships are the building blocks for the test of GARP. Therefore, an incorrect assignment of direct revealed preference in the case of similar bundles may affect the violation rate.

To allow for measurement error, Afriat (1967) proposed removing from the hypothesis test the preference information obtained from the direct comparison of consumption bundles with similar total expenditures (at i's prices) by modifying the direct revealed preference relationship as follows:

$$i\,DRP(e)\,j \text{ if and only if } (e * P_i X_i) \geq P_i X_j \quad (2)$$

where e is a value between zero and one. As e ranges from one to zero, it becomes more difficult to assign the *DRP* relationship. There are several advantages to accounting for measurement error by modifying the direct revealed preference relationship itself. First, an inappropriate assignment of direct preference due to measurement error will not, through a chain of preference, cause GARP violations between completely different consumption bundles.[11] Second, compu-

[10] An alternative would be to prespecify the critical power for the test of GARP and use Monte-Carlo techniques, following Bronars (1987), to determine how similar expenditures must be to attain approximately that power. However, the approximate power of the test is defined in relation to a specified alternative model of household choice and is computationally burdensome for these tests of GARP.

[11] Suppose there is a revealed preference cycle as follows: i *DRP* j, j *DRP* k, k *DRP* l, and l *DRP* i. Consumption bundles i and l will violate GARP. Suppose, however, that j appeared to be *DRP* to k because of measurement error. Under the modified revealed preference rule, j would no longer be *DRP* to k and would no longer "cause" i and l to violate GARP.

tational costs are low. Finally, removing *DRP* relationships established by the comparison of similar consumption bundles may affect both GARP-consistent and GARP-violating comparisons. As a result, accounting for measurement error will not necessarily lead to a decline in the violation rate.[12]

There are two disadvantages to Afriat's proposal. First, it is precisely those pairs of consumption bundles where the difference in total expenditures is small that provide the strongest tests of GARP. Second, as noted by Varian (1990) and Gross (1990), although total expenditures across bundles may be similar, the quantities chosen can be quite different. Therefore, it may be seriously misleading to conclude that a data set is consistent with demand theory and GARP after deleting many of the *DRP* relationships where the difference in total expenditures is small using Afriat's e.

In this study, commodity bundles i and j will be considered similar if both total expenditures and budget shares at i's prices are sufficiently similar. Let d_{kij} be the absolute value of the difference in good k's expenditure share between bundles i and j, evaluated at i's prices. Define consumption bundles i and j to be similar if:

Definition 2: consumption bundles i and j are similar if and only if

(i) expenditure similarity: $e < (P_i X_j / P_i X_i) < (1/e)$; and
(ii) allocative similarity: for goods $k = 1, \ldots, m$ the maximum $d_{kij} < r$.

Consider a definition of revealed preference that accounts for measurement error by assigning no revealed preference relationship to sufficiently similar pairs of consumption bundles. Define the $DRP(e,r)$ relationship as follows:

$$i\ DRP(e,r)\ j \text{ if and only if } P_i X_i \geq P_i X_j \text{ and } i \text{ and } j \text{ are not similar,} \quad (3)$$

where the researcher chooses both r, the critical value for allocative similarity, and e, the critical value for expenditure similarity. The original *DRP* relationship, defined in equation (1), is simply $DRP(1,0)$. Afriat's approach, defined in equation (2), can be expressed as $DRP(e,0)$. Note in equation (3) that as either e decreases or r increases, the number of pairwise comparisons where neither bundle is revealed preferred will rise. These comparisons are not counted as GARP consistent so that the violation rate will not be biased downward when allowing for measurement error in the analysis. Adding the condition that budget shares must be similar solves the problems inherent with Afriat's e alone; (i) *DRP* relationships are removed only when two consumption bundles are actually similar and (ii) consumption bundle comparisons with the strongest possibility of rejecting GARP are not automatically excluded from the hypothesis test.

IV. Non-price Determinants of Demand

The key limitation of any test of optimizing behavior using cross section data is that preference maximization and the existence of a single utility function for a group of consumers are jointly tested. In this case i and j represent different people, not different time periods. In parametric studies, differences in tastes across demographic groups can be introduced by allowing the parameters of the estimated utility function to vary across demographic groups either (i) by specifying that certain of the demand system parameters depend on the demographic variables (see Pollak and Wales (1981)) or (ii) estimating a separate demand system for each demographic group. Option (ii) is appropriate for these nonparametric tests. The goal of this analysis is to determine whether differences in prices and total expenditures alone explain differences in household's consumption choices within a group. Therefore, GARP is tested on relatively homogeneous groups of households.

This approach also provides a method to assess the importance of allowing preferences to differ across demographic groups. If it is important to allow for differences in tastes across demographic groups then the consumption choices of households with different demographic characteristics will be more likely to violate GARP than the consumption choices of households in the same demographic group. Therefore, the violation rate will be computed for a random sample of households across all demographic groups and compared to the violation rate for homogeneous groups of households. If the violation rates are

[12] In the limit, the data will pass GARP because no preference information will be provided by any pairwise comparison.

similar and below the critical value, the hypotheses that *all* households in this cross section have the same preferences and maximize utility will not be rejected.

V. Data

An internal Bureau of Labor Statistics demand analysis data set created by Nelson in 1988 is the data for this study. This data set combined annual consumer unit (CU) expenditures and demographic characteristics from the 1982–85 Consumer Expenditure Interview Survey (CE) and intertemporal, interarea price indices for CUs in twenty-five major metropolitan areas.[13] Annual CU expenditures were aggregated into nine categories: food at home, food away from home, household furnishings, medical care, personal care, transportation, apparel, housing[14] and "other." Intertemporal–interarea price indices for all goods except "other" were created for good i in period t for metropolitan area j as follows:

$$P_{jt}^{i} = \left(CPI_{jt}^{i}/CPI_{j0}^{i}\right) * SI_{j0}^{i},$$

for $i = 1,\ldots,8$ goods; $j = 1,\ldots,25$ areas; and $t =$ the 12 month period the CU was in the survey. CPI_{jt}^{i} is the average annual consumer price index for the 12 month survey period (reference year = Autumn 1973) for good i in city j in time period t. SI_{j0}^{i} is an interarea price index, developed by Sherwood (1975), for good i in area j at period 0 (base year = 1973).[15] The price of "other" was computed as the weighted average sum of the city CPI changes, rebased to 1973, for the four CPI indices that essentially comprise the "other" category in the Nelson data set (entertainment, tobacco, alcohol, and other rental costs) and multiplied by the Sherwood index. Nelson's sample was limited to CUs in the 25 metropolitan areas for which interarea–intertemporal price indices could be computed. This resulted in a sample of 4158 CUs.[16] The CU reports nominal expenditures. Quantity indices are the ratio of expenditures to the intertemporal/interarea price indices described above.

The sample is divided into mutually exclusive demographic groups on the basis of the following eight characteristics: (1) one person households, married households with or without children, and "other" households, (2) whether the reference person is white or nonwhite, (3) whether there are 0, 1, or 2 labor force participants in the CU, (4) whether the reference person is young (under 30 years old), middle-aged (ages 30 to 64), or old (over 64 years old), (5) whether the consumer unit rents or owns it's home,[17] (6) for single and other CU types, the sex of the reference person, (7) the number of CU members, and (8) whether there is a child less than six years old in the CU. Using the above 8 characteristics, there were 43 household groups with at least 20 CUs. These 43 groups account for two-thirds of the data.[18]

VI. Results

In the tests of GARP presented below, it is important to keep several parameters in mind. First, K denotes the cutoff value that determines the number of consumption bundle comparisons used in the calculation of the violation rate. Therefore, changes in K affect the power of the test. Second, the values of e, the critical value for expenditure similarity, and r, the critical value for allocative similarity, modify the *DRP* relationship for measurement error. *DRP*$(1,0)$ tests GARP in the absence of measurement error.

[13] A consumer unit is defined as a group of individuals who share expenditures on two of the following three categories: food, housing, and all other.

[14] Housing is measured on a rental equivalence basis for home owners.

[15] CPI_{jt}^{i} are appropriate measures of the change in prices over time for a given city but cannot be used to compare changes in prices across cities. The Sherwood indexes are the only appropriate interarea measure of prices available and have been used in several research studies. However, the weights for the Sherwood indices are based upon a 1960s basket. The prices for the basket are based in 1973. As such, updating these indices to the 1980s is less than ideal, though the change in the weights *within* the nine categories may not be that great.

[16] Nelson excluded CUs if they changed rental/homeownership status during the year, if their demographic composition was unstable, or if they moved 25 miles or more. I excluded CUs if they had imputed earnings or hours data, negative medical expenditures, or missing homeownership status.

[17] The housing price and expenditure data for homeowners are based on rental equivalence and may be subject to more measurement error than the housing data for renters. As a result, tests of GARP are performed on homeowners and renters separately.

[18] Sample statistics, descriptions of the 43 groups, and descriptions of the price, expenditure and quantity variables are available from the author.

TABLE 1.—GARP TESTS FOR *DRP*(1, 0): AVERAGE RESULTS FOR 42 DEMOGRAPHIC GROUPS

Variable	Median	Mean (Std. Dev.)	Min	Max
Number of Observations	43.50	60.52 (42.69)	20.00	180.00
Pairwise Comparisons/100	18.49	53.81 (76.10)	3.80	322.20
Number *DRP*/100	9.11	26.68 (37.75)	1.87	159.02
GARP Consistent/100	18.17	52.32 (73.24)	3.68	307.79
Neither *i RP j* Nor *j RP i*	20.00	25.33 (22.86)	0.00	86.00
GARP Violations	12.00	123.60 (293.20)	0.00	1379.00
Violation Rate * 100	0.70	1.10 (1.11)	0.00	4.29

Violation Rates Calculated Separately Among Pairwise Comparisons with Similar Total Expenditures

	Violation Rate		Number of Comparisons	
K	Median	Mean (Std. Dev.)	Median	Mean (Std. Dev.)
$K \leq 0.05$	5.13	6.40 (6.35)	168.00	413.62 (623.11)
$K \leq 0.10$	3.87	5.37 (4.97)	309.00	818.76 (1235.01)
$K \leq 0.15$	3.16	4.33 (3.92)	485.00	1222.14 (1848.24)
$K \leq 0.20$	2.67	3.54 (3.31)	621.00	1631.48 (2428.49)

A. *Tests of GARP*

In test of GARP using *DRP*(1, 0), one of the 43 household types was clearly an outlier and is discussed below. A summary of the tests of GARP for the remaining 42 demographic groups is presented in table 1. Only four of the groups had consumption choices and budget constraints that were consistent with GARP.[19] The median demographic group had 12 pairs of consumption choices that violated GARP. Is it reasonable to estimate a parametric model for these 42 demographic groups? As argued earlier, the number of pairwise comparisons that are consistent with GARP can be used to place the number of violations into context. The median demographic group had 1817 GARP-consistent comparisons and so a violation rate of 0.7%.

The violation rate presented above does not take into account that some pairs of consumption choices are unlikely to ever violate GARP, regardless of how choices are made. To diminish the impact of pairwise comparisons that are trivially consistent with GARP, the violation rate was calculated among pairwise comparisons of consumption bundles with similar expenditures (small values of K).[20] The bottom panel of table 2 presents these results. As expected, the violation rate rises as it is computed across consumption bundles with more similar total expenditures (in base period prices): the median violation rates are 2.67%, 3.16%, 3.87%, and 5.13% for households where K is 0.20, 0.15, 0.10, and 0.05, respectively. However, even within narrow total expenditure ranges, there are *many* more comparisons that fail to reject GARP than there are violating comparisons. Of the 42 demographic groups, 33 had violation rates that were less than 5% among pairwise comparisons where (constant price) expenditures differed by no more than 20%.

In these data, GARP violations *primarily* occur among consumption bundles with similar total expenditures. Averaging across the 42 groups, 72% of all GARP violations occur among consumption bundle comparisons with $K \leq 0.10$. In fact, 94% of all GARP-violating pairs have $K \leq 0.20$. On the other hand, 72% of GARP consistent pairs have $K > 0.20$. Therefore, calculating the violation rate among pairwise comparisons with similar total expenditures increases the power of these tests of GARP.

White, two person, homeowners with a family head over the age of 65, where neither person participated in the labor market, is the outlier

[19] All are married couples: (i) four person, middle-aged head, nonwhite, homeowner, dual earners, (ii) two person, old family head, white, homeowner, dual earners, (iii) two person, old family head, white, homeowner, single earner, and (iv) two person, young head, white, renters, dual earners.

[20] The price vector for Seattle, May 1984, is the base period price vector (P_1) used to compute M_i.

household type. The *overall* violation rate for this demographic group is over 11.4%. The highest violation rate overall among the remaining 42 groups is 4.3%. The violation rate among consumption bundle comparisons where K equals 0.20, 0.15, 0.10, and 0.05 is 28.5%, 30.1%, 32.7%, and 34.2%, respectively, for this older homeowner sample.

There are several plausible hypotheses for the inconsistency of this group with demand theory. First, the medical expenditures budget share is much higher for this group of households than on average: 10.3% versus 5%. Large, unplanned medical expenditures may lead to GARP violations. In addition, health status may affect preferences. Unfortunately, there is no direct information about the health of the household in the CE so it is not possible to test this hypothesis. Housing's budget share is also greater for this group: 42% versus 35% on average. It may be more costly for older homeowners to adjust their housing consumption. Therefore, the problems that arise for the static utility model if there is lagged adjustment may be particularly important for this group. Finally, with 222 households, this demographic group is the largest in the sample. More consumption bundle comparisons with similar total expenditures increase the likelihood of GARP violations because there are more ways for one consumption bundle to be revealed preferred to another. Section D examines this hypothesis. The remaining summary statistics for the tests of GARP exclude this outlier demographic group.

B. *GARP Results Accounting for Measurement / Optimization Errors*

The hypothesis of common preferences and utility maximization works quite well for the majority of these demographic groups. Nevertheless, only four household types had consumption choices strictly consistent with revealed preference axioms. One hypothesis is that measurement error contributes to violations of GARP. To examine this hypothesis, tests of GARP are performed using the modified definition of *DRP* proposed in equation (3).

Table 3 presents the results of tests of GARP using $DRP(0.99, 0.10)$. There are 911 $DRP(1, 0)$ relationships on average per demographic group and, on average, only *one* $DRP(1, 0)$ relationship per subsample satisfied both the inequality constraints on expenditure similarity and allocative similarity in definition 2. Even though only one of the $DRP(1, 0)$ relationships is removed on average, violation rates fall somewhat when specifying $DRP(0.99, 0.10)$. For K equal to 0.2, 0.15, 0.10, and 0.05 violation rates are 13.8%, 20.8%, 10%, and 6.6% *lower* when allowing for measurement error. Though not reported in a table, tests of GARP were also conducted for $DRP(0.95, 0.05)$. On average 5 of the $DRP(1, 0)$ relationships per subsample are removed. However, each *DRP* relationship that was removed was replaced by an *RP* preference ordering. Therefore, GARP test results are identical for $DRP(0.95, 0.05)$ and $DRP(1, 0)$.

C. *Examining the Power of the Tests of GARP*

Household's consumption choices and budget constraints in these data seldom violate revealed preference axioms, even when total expenditures are similar. However, the violations that do occur are troubling because most GARP violations compare quite different consumption bundles. A possible explanation for this finding is that there is too little price variation in these data. If budget lines rarely intersect then, even when total expenditures are similar, the power of tests of GARP is low (see Bronars (1987)). To provide some measure of the power of these tests of GARP, I assigned a randomly selected price vector to each household's consumption vector and retested GARP. If the power of the test is low, then violation rates will not increase when using the randomly assigned prices.

For each of the 42 demographic groups, price vectors were randomly assigned 100 times and tests of GARP were performed. Across the 4,200 tests of GARP, median violation rates increased from 2.67% to 6.8% for households with expenditures within 20% and increased from to 3.87% to 12.3% for households with expenditures within 10%. Median violation rates among household comparisons with similar expenditures more than double when prices are randomly assigned. In addition, the joint hypotheses of preference maximization and common preferences would be rejected for both $K = 0.20$ and $K = 0.10$ on the basis of tests of GARP with randomly assigned prices. Given the substantial increase in median

TABLE 2.—ALLOWING FOR MEASUREMENT ERROR: GARP TESTS FOR *DRP*(0.99, 0.10) (AVERAGE RESULTS FOR 42 DEMOGRAPHIC GROUPS)

Variable	Median	Mean (Std. Dev.)	Min	Max
GARP Consistent/100	18.14	52.38 (73.48)	3.72	308.48
GARP Violations	10.00	111.05 (262.25)	0.00	1288.00
Neither *i RP j* Nor *j RP i*	22.00	31.76 (30.76)	0.00	138.00
Violation Rate * 100	0.68	0.98 (0.96)	0.00	4.01
Dropped *DRP* Rate (per 100 GARP *DRP*)	0.27	0.26 (0.21)	0.00	1.05

Violation Rates Calculated Separately among Pairwise Comparisons with Similar Total Expenditures

K	Median	Mean (Std. Dev.)
$K \leq 0.05$	4.47	5.83 (5.67)
$K \leq 0.10$	3.14	4.81 (4.39)
$K \leq 0.15$	2.86	3.91 (3.53)
$K \leq 0.20$	2.50	3.20 (2.98)

TABLE 3.—GARP TESTS FOR *DRP*(1, 0): AVERAGE RESULTS FOR 500 RANDOM GROUPS OF HOUSEHOLDS

Variable	Median	Mean (Std. Dev.)	Min	Max
Number of Observations	56.5	79.93 (58.97)	24.00	269.00
Pairwise Comparisons/100	31.36	97.79 (148.05)	5.52	720.92
GARP Consistent/100	30.77	95.52 (143.37)	5.40	707.88
GARP Violations	22.00	190.60 (524.43)	0.00	5867.00
Neither *i RP j* Nor *j RP i*	22.00	35.76 (34.92)	0.00	244.00
Violation Rate * 100	0.76	1.00 (9.03)	0.00	8.15

Violation Rates Calculated Separately among Pairwise Comparisons with Similar Total Expenditures

	Violation Rate		Number of Comparisons	
K	Median	Mean (Std. Dev.)	Median	Mean (Std. Dev.)
$K \leq 0.05$	7.38	8.21 (6.02)	160.00	421.83 (585.72)
$K \leq 0.10$	5.93	6.99 (5.10)	300.00	841.77 (1168.06)
$K \leq 0.15$	4.88	5.69 (4.36)	445.00	1255.42 (1738.98)
$K \leq 0.20$	3.82	4.69 (3.73)	592.00	1667.12 (2309.49)

GARP violation rates among households with similar expenditures when prices are randomly assigned, it appears that the tests of GARP in the previous section are fairly powerful.

D. *Determining the Importance of Subsetting by Demographic Group*

In this section, I assess the importance of grouping households by demographic type. Tests of GARP using *DRP*(1, 0) were performed on 500 random samples of households. If the consumption choices of households chosen at random are as consistent with revealed preference axioms as the choices of households with similar demographic characteristics, the joint hypotheses that households maximize utility and that *all* households have the same preferences will not be rejected.

The 500 random samples were chosen so that the distribution of consumption bundle comparisons with $K = 0.10$ approximates the distribution across the 42 demographic groups. It is important that the *number* of comparisons with similar total expenditures be approximately the same when comparing violation rates across samples. The greater the number of comparisons with similar total expenditures, the greater the likelihood that GARP will be violated because there are more ways for one bundle to be *RP* to another.[21]

Table 3 presents the results of the tests of GARP for the 500 randomly assigned household groups. Comparing the bottom panels of tables 1 and 3, the mean and median number of consumption bundle comparisons where K is 0.20, 0.15, 0.10, and 0.05 are quite similar across the 42 demographic subsamples and 500 randomly assigned household groups (by construction). Therefore, it is valid to compare violation rates among households with similar total expenditures. For $K = 0.20$ and $K = 0.10$, the median violation rate is 43% and 53% higher, respectively, across the 500 randomly assigned groups than across the 42 demographic groups. In fact, the median violation rate across the 500 groups of households of all demographic types is more than 5% for values of $K \leq 0.10$. Therefore, the hypothesis that all households have the same set of preferences and maximize utility may be rejected for the 500 randomly assigned groups, for reasonable values of K. Allowing for preference heterogeneity across demographic groups does improve consistency with GARP.

Violation rates were averaged across 10 samples of 270 households chosen at random to determine whether the high violation rate among the older, white homeowners demographic subsample is simply the result of its large sample size. The median violation rates across the 10 randomly assigned samples for consumption bundle comparisons with K equal to 0.20, 0.15, 0.10, or 0.05 is 14.4%, 16.8%, 19.8%, and 22.0%, respectively. Violation rates in these large random samples of households are approximately 50% *lower* than in the married, white, homeowner sample. Therefore, the large sample size alone does not account for the inconsistency of this demographic group.

[21] The variation in total expenditures for a household group of size N is greater among a randomly chosen group than among households with similar demographic characteristics. Therefore, sample sizes in the randomly assigned household groups are approximately 22% larger than in the 42 groups with the same demographic characteristics so that the number of comparisons with $K = 0.10$ will be similar.

VII. Conclusions

This paper has provided the first nonparametric tests of demand theory using household microeconomic data. The actual consumption choices of decision-making consumer units were tested for consistency with revealed preference axioms. In addition, two methodological issues in performing tests of GARP were addressed. The violation rate, or the fraction of consumption bundle comparisons that violate GARP, was proposed as a summary statistic for tests of GARP. To increase the power of the test, the violation rate was calculated separately among the consumption bundle comparisons that have similar total expenditures (in constant dollars). Finally, a modification of Afriat's Efficiency Index was proposed as a method to account for measurement error.

There are several important findings in this study. First, households with similar demographic characteristics make consumption choices that are generally consistent with GARP. Second, allowing for measurement error reduces the violation rate somewhat. Finally, the consumption choices and budget constraints of households with different demographic characteristics were generally not consistent with GARP. Therefore, allowing preferences to differ across demographic groups increases the consistency of observed choices with demand theory. For these data, estimation of a parametric, static utility model that allows for preference heterogeneity is warranted by the nonparametric results.

REFERENCES

Afriat, Sydney N., "The Construction of Utility Functions from Expenditure Data," *International Economic Review* 8 (1967), 67–77.

Banker, Rajiv D., and Ajay Maindiratta, "Nonparametric Analysis of Technical and Allocative Efficiencies in Production," *Econometrica* 56 (6) (1988), 1315–1332.

Becker, Gary S., "Irrational Behavior and Economic Theory," *Journal of Political Economy* 70 (1962), 1–13.

Blundell, Richard, Panos Pashardes, and Guglielmo Weber, "What Do We Learn about Consumer Demand Patterns from Micro Data?" *American Economic Review* 83 (3) (1993), 570–597.

Bronars, Stephen G., "The Power of Nonparametric Tests of Preference Maximization," *Econometrica* 55 (3) (1987), 693–698.

Deaton, Angus, "Demand Analysis," in Zvi Griliches and Michael D. Intriligator (eds.), *Handbook of Economet-*

rics, Volume III (Amsterdam: Elsevier Sciences Publishers BV, 1986), 1767–1839.

Diewert, W. Erwin, and Celic Parkan, "Tests for the Consistency of Consumer Data," *Journal of Econometrics* 30 (1985), 127–147.

Gross, John, "On Expenditure Indices in Revealed Preference Tests," *Journal of Political Economy* 99 (2) (1990), 416–419.

Houtman, Martijn, and J. A. H. Maks, "Determining All Maximal Subsets Consistent with Revealed Preference," *Kwantitatieve Methoden* 19 (1985), 89–104.

Landsburg, Steven E., "Taste Change in the United Kingdom, 1900–1955," *Journal of Political Economy* 89 (1981), 92–104.

Manser, Marilyn E., and Richard McDonald, "An Analysis of Substitution Bias in Measuring Inflation," *Econometrica* 56 (1988), 909–930.

Pollak, Robert A., and Terence J. Wales, "Demographic Variables in Demand Analysis," *Econometrica* 49 (1981), 1533–1551.

Swofford, James L., and Gerald A. Whitney, "Nonparametric Tests of Utility Maximization and Weak Separability for Consumption, Leisure, and Money," this REVIEW 69 (1987), 458–464.

Varian, Hal R., "The Non Parametric Approach to Demand Analysis," *Econometrica* 50 (1982), 945–973.

______, "Non-Parametric Analysis of Optimizing Behavior with Measurement Error," *Journal of Econometrics* 30 (1985), 445–458.

______, "Goodness-of-Fit in Optimizing Models," *Journal of Econometrics* 46 (1990), 125–140.