Macroeconomic Dynamics, **4**, 2000, 324–342. Printed in the United States of America.

GARP, SEPARABILITY, AND THE REPRESENTATIVE AGENT

ADRIAN R. FLEISSIG
California State University, Fullerton

ALASTAIR R. HALL AND JOHN J. SEATER
North Carolina State University

We examine whether annual, quarterly, and monthly U.S. aggregate consumption data could have been generated by a utility-maximizing representative agent with intertemporally separable utility. The model appears inapplicable over the full time periods covered by the NIPA data, which are the sample periods often used in the literature. The model does appear applicable, however, over long subsamples. The data also are inconsistent with separability assumptions routinely made in the literature. In particular, the main categories of consumption (nondurables, services, and durables) are not mutually separable. We consider the implications of our results for inference about consumption based on the representative-agent model.

Keywords: Nonparametric, Utility Maximization, Consumption Data

> "To assume that the representative agent acts like an ideal consumer is a hypothesis worth testing: to assume that an actual person, the Mr. Brown or Mr. Jones who lives round the corner, does in fact act in such a way does not deserve a moment's consideration." Hicks (1956, p. 55)

1. INTRODUCTION

The representative-agent model is a popular approach to analyzing aggregate consumption. Its basic assumption is that one can treat the aggregate data as the outcome of the decisions of a single "representative" consumer. However, the conditions on individual preferences necessary for the representative agent to be an exact representation of the behavior of underlying agents are quite stringent, so much so as to be implausible. Most criticism of the representative-agent paradigm arises from this problem; for example, see Kirman (1992). Nonetheless, the representative-agent model can be useful even without a complete microfoundation. A central task of macroeconomics is to explain aggregate data and make predictions. The representative-agent model is useful if it helps to do so. Exact aggregation of individual preferences is not essential. What is essential is that there

We thank Bill Barnett, an associate editor, Douglas Fisher, Ronald Gallant, Philip Meguire, Douglas Pearce, James Swofford, and David Wilcox for helpful comments. We are extremely grateful to Hal Varian for providing us with his NONPAR software to perform the GARP and separability tests. Address correspondence to: Adrian R. Fleissig, Department of Economics, California State University, Fullerton, Fullerton, CA 92834, USA; e-mail: afleissig@fullerton.edu.

© 2000 Cambridge University Press 1365-1005/00 $9.50

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

exists some "representative" utility function consistent with the data, for only then can the researcher resort to standard utility theory in conducting the analysis.

The usual approach is to assume that a representative utility function exists and that it satisfies certain separability restrictions, permitting analysis of convenient subsets of the data. Empirical results then are interpreted in terms of individual behavior, and any statistical rejection of the model is treated as evidence of the misspecification of the representative agent's utility function. See, for example, the recent survey by Kocherlakota (1996). Another possibility, however, is that no representative utility function with the implied separability structure exists. Then, subsequent interpretations of results in terms of individual behavior are suspect at best, and empirical rejections might reflect invalidity of the entire representative-agent approach rather than merely misspecification of that agent's utility function.

One can check whether a representative-agent utility function exists by using Varian's (1982, 1983) tests of whether a set of data is consistent with the generalized axiom of revealed preference (GARP). Consistency with GARP is necessary and sufficient for existence of a well-behaved utility function capable of generating the data. An important advantage of Varian's approach is that the tests are nonparametric and so require only mild assumptions on the form of the utility function.

We use Varian's tests to see if a representative-agent utility function exists for the monthly, quarterly, and annual aggregate consumption data of the United States.[1] The tests require user costs for the durable goods, which are unavailable in official statistics. We therefore construct measures ourselves, which requires assumptions about the expectation mechanism of the representative agent. We tried several expectations models, but the results are not very sensitive to which of those is used. We start by testing whether the data are consistent with a well-behaved intertemporally separable utility function. We find that the data generally are not consistent with such a function over the full periods for which the data are available. Those are the sample periods often used in previous empirical work. However, we can identify the periods responsible for the violations and thus uncover subsamples of the data over which GARP is satisfied. We find that there are long subsamples of both the annual and quarterly data that are consistent with GARP, and the subsamples for the two kinds of data are essentially the same. In particular, 1981 is the only year of instability for both the annual and quarterly data. In contrast, the monthly data show a great many violations of GARP. The subsamples of monthly data consistent with GARP are so short that estimation of the utility function is likely to be impractical.

Having identified periods over which GARP holds, we explore the structure of the representative utility function within these periods. We apply Varian's (1982, 1983) separability tests to several possible groupings of aggregate consumption's components, such as furniture, clothing, automobiles, and so forth. Again, the annual and quarterly data yield very similar results. Our results support the analysis of nondurables and services as a group without the inclusion of durables. However, if all three categories are analyzed at once, then durables must be divided into

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

certain subgroups. Interestingly, our results indicate that automobiles should be treated as a separate subgroup, as recently advocated by Wilcox (1992) after his analysis of the survey methods by which the data are collected.

An outline of the paper is as follows. Section 2 describes the theory behind Varian's tests and discusses the properties of the tests. Section 3 describes the data used in our study and in particular the calculations needed to construct price variables for durable goods. Section 4 contains the results and Section 5 concludes with a discussion of some implications of our results.

2. TESTING GARP AND THE SEPARABILITY STRUCTURE OF THE UTILITY FUNCTION

The standard representative-agent model supposes that the agent maximizes his lifetime utility,

$$E_0\left[\sum_{t=0}^{T} u(\boldsymbol{x}_t)(1+\rho)^{-t}\right] \tag{1}$$

subject to his lifetime budget constraint

$$\sum_{t=0}^{T} \boldsymbol{p}_t \boldsymbol{x}_t (1+r)^{-t} = A_0 + \sum_{t=0}^{T} y_t (1+r)^{-t}, \tag{2}$$

where u is the aggregate intratemporal utility function, $\boldsymbol{x}_t$ is a k-vector of goods $\{x_{it}\}$, $\boldsymbol{p}_t$ is a corresponding vector of relative prices, A_0 is initial financial assets, y_t is real labor income (here taken as exogenous, for simplicity), r is the real interest rate, ρ is the rate of time preference, T is the time horizon, and $E_0[\cdot]$ denotes expectations conditional on information at time 0. Possible sources of uncertainty include y, r, and T. Notice that the representative agent's utility function is assumed to be intertemporally separable. This restriction is imposed in many of the models underpinning macroeconomic analysis of consumption. However, it rules out such features as habit formation; we return to this issue in Section 5.

Whether the foregoing model is appropriate for analyzing aggregate consumption data is an empirical question that can be addressed with Varian's (1982, 1983) procedures for testing consistency with GARP. We therefore begin with a brief review of those tests and their background.

2.1. Testing GARP

GARP is simply a mild generalization of the strong axiom of revealed preference (SARP) that allows "flat spots" in the preference relation, so that the marginal rate of substitution can be constant along a section of the indifference curve. Precise statements of SARP and GARP are given in Appendix A. GARP's importance here is that we can use it to see if there is a utility function that could have generated a given set of data.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

Suppose that we have T observations on the k-vector $\boldsymbol{x}_t$ of consumption goods and on its corresponding k-vector of prices, $\boldsymbol{p}_t$. The following definition is from Varian (1983).

DEFINITION 1. *Rationalizing Utility Function*: *A utility function u rationalizes the data* $\{\boldsymbol{x}_t, \boldsymbol{p}_t\}$, $t = 1, \ldots, T$, *if* $u(\boldsymbol{x}_t) \geq u(z)$ *for all z such that* $\boldsymbol{p}_t \boldsymbol{x}_t \geq \boldsymbol{p}_t z$ *for all t*.

This definition simply means that $u(\cdot)$ is consistent with the data if observed consumption would be optimal under $u(\cdot)$. Varian (1982) develops methods for examining whether any such utility function exists for a given data set based on the following theorem due to Afriat (1967) and Diewert (1973):

AFRIAT'S THEOREM. *The following conditions are equivalent*:

1. *There exist numbers* U_t, $\lambda_t > 0$ *that satisfy the Afriat inequalities*

$$U_t \leq U_s + \lambda_s p_s (x_t - x_s)$$

   for $s, t = 1, \ldots, T$.
2. *The data satisfy GARP*.
3. *There exists a nonsatiated utility function that rationalizes the data.*
4. *There exists a concave, monotonic, continuous, nonsatiated utility function that rationalizes the data*.

By parts 3 and 4 of the theorem, if there is a utility function that rationalizes the data, it will have the properties typically assumed in consumer theory. By parts 2 and 3, some rationalizing utility function exists if and only if the data satisfy GARP. By parts 1 and 2, we can test for GARP by seeing if there is a set of numbers $\{U, \lambda\}$ satisfying a set of linear inequalities. It is a straightforward linear programming exercise to see if an appropriate set $\{U, \lambda\}$ exists.[2] That exercise provides the method for testing whether the data satisfy GARP: simply check if the Afriat numbers $\{U, L\}$ exist. Note that Afriat's theorem says that GARP is equivalent to the existence of a rationalizing utility function. Therefore, if Afriat numbers cannot be found for the data, a rationalizing utility function does not exist.

2.2. Testing Separability Assumptions

Suppose the consumption data satisfy GARP. We then know that there is some utility function $u(x)$ that rationalizes the data. However, that function has as its argument the entire k-vector of consumption goods x, and analysis of consumption behavior is inconvenient or even intractable if k is large. We can solve this problem by aggregating the elements of x into various subgroups of expenditures.[3] The question is which elements of x can be combined legitimately into a particular group. That questions also can be addressed with GARP tests. To see how, we need another definition and another theorem.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

Suppose the vector $\boldsymbol{x}$ is partitioned into two sets $a = (x_1, \ldots, x_m)$ and $b = (x_{m+1}, \ldots, x_k)$ with corresponding price vectors $\boldsymbol{p}_a$ and $\boldsymbol{p}_b$. We then have the following definition.

DEFINITION 2. *Weak Separability*: *A utility function* $u(x)$ *is weakly separable in the b goods if there exists a subutility function* $v(b)$ *and a macro function* $u^*(a, v(b))$ *that is continuous and monotonically strictly increasing in* $v(b)$, *such that* $u(a, b) \equiv u^*(a, v(b))$.

The subutility function v is a utility function in its own right with the usual properties. Weak separability of u in the b goods means that the marginal rate of substitution between any two elements of the b vector is independent of all elements of the a vector. Note that the utility function can be separable in b but not in a.[4]

When the utility function u is separable in b, its value does not depend on the elements of b individually but rather on their combined subutility value $v(b)$. The household in effect solves its budgeting decision in two stages: First, it divides expenditures between a and b; second, it decides the optimal allocation among the elements of b independently of the choices of the elements of a (Gorman, 1959). When utility is separable, the econometrician can confine analysis to the b elements and simply disregard the a goods (except in specifying the relevant budget constraint). In particular, if the utility function is separable in nondurables and services, one can ignore the service flow of durables in studying the household's choice between nondurables and services, greatly simplifying the analysis.[5]

Varian (1982, 1983) provides two tests of separability. The first test is the weaker of the two, checking if necessary consistency conditions are met. The subdata b must satisfy GARP because each observation on b must solve the problem

$$\begin{aligned} &\max v(b) \\ &\text{subject to} \quad p_{bt} b \leq p_{bt} b_t . \end{aligned}$$

So it is necessary for separability that Afriat numbers exist for the b subdata because, otherwise, the v subutility function does not exist. A test of necessary conditions therefore can be done by checking if the subdata b satisfy GARP.

The second test is stronger because it checks necessary and sufficient conditions for separability. For sufficiency, the Afriat numbers for the subutility function v must be consistent with those for the utility function u. The following theorem from Varian (1983) states the relationship precisely:

VARIAN'S THEOREM. *The following conditions are equivalent*:

(i) *There exist numbers* $U_t, V_t, \lambda_t, \mu_t > 0$, *which satisfy the Afriat inequalities*

$$\begin{aligned} U_t &\leq U_s + \lambda_s p_{as}(a_t - a_s) + (\lambda_s/\mu_s)(V_t - V_s), \\ V_t &\leq V_s + \mu_s p_{bs}(b_t - b_s) \end{aligned}$$

for $s, t = 1, \ldots, T$.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

(ii) *The data* (p_{bt}, b_t) *and* $(p_{at}, 1/\mu_t; a_t, V_t)$ *satisfy GARP for some choice of* (V_t, μ_t) *that satisfies the Afriat inequalities.*

(iii) *There exists a weakly separable nonsatiated, continuous, concave, monotonic utility function that rationalizes the data.*

Part (i) provides the means for testing the necessary and sufficient conditions; one must construct two sets of interrelated Afriat numbers. Varian has written a computer routine to do the calculations. However, Varian's program actually performs only a sufficient test for weak separability because of computational constraints; see Swofford and Whitney (1994).[6] Therefore, if the data pass the necessary conditions in the first test but fail the sufficient conditions in the second test, we cannot rule out the possibility that the separability assumption in question in fact does hold. Of course, if the necessary condition fails or the sufficient condition passes, then the results of the tests are definite.

2.3. Test Implementation and Properties

The GARP and separability tests are done with Varian's NONPAR software, which he kindly provided. NONPAR does the linear programming exercise necessary to see if the data satisfy GARP and reports any violations. A violation occurs when a consumption bundle is revealed as not preferred in one period but then is bought in another period even though the bundle previously revealed as preferred is a feasible choice.

The NONPAR tests have both advantages and disadvantages. The main advantage is that the tests are nonparametric; one need not specify the form of the utility function. Also, the tests can handle a large number of goods. The main disadvantage is that the tests are nonstochastic. Violations are all or nothing; either there is a utility function that rationalizes the data or there is not. We therefore must be especially careful about the possibility of measurement error. If there were no measurement error, then any observed rejection of GARP would be a genuine rejection. However, with measurement error, false rejections may occur, and without a distribution theory for the tests, we cannot judge the importance of observed rejections by use of the conventional significance tests.[7] We respond to this difficulty by faithfully reporting all rejections and by adding some judgmental discussion of the likely significance of the test results. The reader should bear in mind, though, that there is a bias toward rejection (Barnett and Choi, 1989) because even one inconsistency leads to total rejection of GARP; consequently, passing the GARP tests is a very strong result.

Despite their problems, the GARP tests provide useful information. It is true that one must exercise caution in interpreting the tests because of their nonstochastic nature. However, the only available alternative is to do no testing and simply proceed with inference conditional on untested assumptions about the existence of a representative agent and the separability properties of his utility function. We believe it preferable to use the GARP tests, particularly in view of the results that we report later.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

3. DATA AND THE USER COST OF DURABLE GOODS

Our goal is to see if there is a utility function consistent with the U.S. aggregate consumption data and if it is separable in any of its arguments. To do this, we perform Varian's tests on monthly, quarterly, and annual data for the United States, spanning the periods 1959:01–1990:12, 1959.1–1990.4, and 1929–1990, respectively. All data are described in Appendix B. We assume that the service flows from durables are proportional to the durable stocks, and so, the flows can be measured by the stocks, as in Deaton and Muellbauer (1980).[8]

The tests require a price for each good. Prices for nondurables and services are readily available in the official statistics, but proper prices for durables are not. We need the user cost of durables, but official statistics report only the sale prices for new durables. We therefore construct measures of the user cost, and we describe how we did so in the remainder of this section.[9]

The formula for the user cost of durable good i in period t is

$$p^{\text{uc}}_{\text{rel},it} = p^{d}_{\text{rel},it} - (1-\delta_i)E_t\left(m_{t,t+1}p^{d}_{\text{rel},it+1}\right), \tag{3}$$

where $p^{\text{uc}}_{\text{rel}}$ is the user cost relative to some nondurable numeraire, p^{d}_{rel} is the relative price of a unit of the durable, δ is the depreciation rate, $m_{t,t+1}$ is the intertemporal marginal rate of substitution for the numeraire consumption good, and $E_t X_{t+1}$ denotes the conditional expectation at time t of X at time $t+1$. The usual practice is to assume that δ is the same for all goods.[10] We do not follow that practice but instead use different depreciation rates for each of the eleven categories of durables. Depreciation rates differ substantially across types of goods, and so, it seems prudent not to impose one average rate on everything.

We must adjust (3) in two ways to use it in constructing the user costs that we need for our tests. First, we cannot observe $m_{t,t+1}$, which depends on the particulars of the utility function. We therefore approximate it by $(1+r_t)^{-1}$, where r is the real interest rate. We thus have

$$p^{\text{uc}}_{\text{rel},it} = p^{d}_{\text{rel},it} - (1-\delta_i)(1+r_t)^{-1}E_t\left(p^{d}_{\text{rel},it+1}\right).$$

This approximation makes no risk correction and thus ignores the effect of uncertainty on the value of assets. Second, the foregoing equation is in terms of relative prices, whereas we need prices in money terms for the GARP tests (because we must multiply each quantity by its price in checking for violations of the GARP conditions). We therefore replace all prices in the right side with their nominal equivalents; that is, money prices p^d replace p^{d}_{rel} and the nominal interest rate R replaces the real rate r. We have, finally,

$$p^{\text{uc}}_{it} = p^{d}_{it} - (1-\delta_i)(1+R_t)^{-1}E_t\left(p^{d}_{it+1}\right). \tag{4}$$

Next, we face the problem of measuring expected prices because the right side of (4) contains the expected durables price $E_t p^{d}_{it+1}$. Previous work, such as that of

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

Diewert (1974) and Swofford and Whitney (1987, 1988), has used static expectations, according to which the expected price for next period is simply the current period's actual price:

$$E_t p^d_{it+1} = p^d_{it}$$

This scheme implies that people think that p^d is a driftless random walk and so do not adjust the current price in predicting next period's price. Inflation implicitly is regarded as white noise, and its expected value always is zero. Such rigid expectations seem implausible, especially because p^d in fact is not a driftless random walk and inflation is not white noise. However, we use the static scheme as one of our models of expectations formation both because it has been used by others and because it provides a kind of lower bound on expectations behavior. At the other extreme is the perfect-foresight model, according to which people know next period's price exactly:

$$E_t p^d_{it+1} = p^d_{it+1}.$$

Perfect foresight, too, seems implausible, and in addition it yields a few observations of negative user costs that required somewhat arbitrary adjustments to correct, as described momentarily.[11] Nonetheless, because of its simplicity, we use it to construct an alternate user cost series.

Given the limitations of both static expectations and perfect foresight, we also estimated ARIMA models for the various p^d_{it} and took expectations of them. We fitted several models to each of the individual durables series but the resulting user costs yielded test results much like those from the perfect-foresight model and provided no additional insights. To save space, we therefore do not report any results based on the ARIMA models.

The annual data include the Second World War, which created a special problem with perfect foresight. Under perfect foresight, the user costs for autos, other autos, furniture, kitchen appliances, other household appliances, video and audio and ophthalmic products are negative for some war years 1941–1946. In those years, there were restrictions on new investment, and repairs sometimes could not be made because of parts and labor shortages. Consequently, there were increases in depreciation and discards not captured by our depreciation-rate measures, which are constant for all years. To adjust for these war-year factors, we interpolated between the 1940 and 1947 user cost levels by repeatedly increasing the depreciation rate for each good by 0.05% per iteration until the user cost was within 15% of the 1940 and 1947 levels, respectively.

4. TEST RESULTS

We now report the results of applying Varian's GARP and separability tests to this data set on aggregate consumption for the United States.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

Table 1. GARP test results

	Data frequency		
Expectations model	Monthly 1959.01–1990.12	Quarterly 1959.1–1990.4	Annual 1929–1990
Static expectations	221	43	0
Perfect foresight	157	55	11

Table 2. GARP-consistent subsample periods

	Data frequency		
Expectations model	Monthly 1959.01–1990.12	Quarterly 1959.1–1990.4	Annual 1929–1990
Static expectations	1960.01–1974.09	1960.1–1990.4	1929–1990
	1974.10–1978.11		
	1978.12–1980.01		
	1980.02–1981.04		
	1981.05–1988.02		
	1988.03–1990.06		
	1990.07–1990.12		
Perfect foresight	1960.01–1968.07	1960.1–1980.4	1935–1981
	1968.07–1970.05	1981.4–1990.4	1982–1990
	1970.05–1990.12		

4.1. GARP Tests

Under the static expectations model, the annual data pass GARP, but the quarterly and monthly data both fail. Under both perfect foresight, all three data sets fail. Table 1 reports the numbers of violations.

Previous tests of GARP in the literature have focused exclusively on the number of violations. However, it is also of interest to know when these violations occur because then we can define subsamples over which GARP is satisfied. We used the following method, apparently original to this paper, to identify these breakpoints. First, we applied the GARP tests to the whole sample. The tests compare every year's consumption bundle and prices to every other year's bundle and prices and report any violations of GARP. For example, if observations 25 and 35 are inconsistent with each other, the tests report violations for observations 25–35 and 35–25. We then split the sample at these points and reran the GARP tests within the resulting subsamples. Working through the violation list this way allowed us to construct the largest possible subsamples consistent with GARP and identify the periods where there is a break in consistency. Table 2 reports the resulting GARP-consistent subsamples.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

With static expectations, the annual data pass GARP over the full sample, the quarterly data pass over the subsample 1960.1–1991.4, and the monthly data have no subsample of any appreciable length that passes. With perfect foresight, the annual data pass GARP over the two subsamples 1935–1981 and 1982–1990, the quarterly data pass over the two subsamples 1960.1–1980.4 and 1981.4–1991.4 (notice the discontinuity), and the monthly data pass over the subsamples 1960.01–1968.07, 1968.08–1970.05, and 1970.06–1993.05. Clearly, the results are sensitive to the method of expectation formation. There are several other noteworthy aspects of these results.

First, for the quarterly and monthly data, the full samples are inconsistent with the decisions of a representative agent whose utility function is intertemporally separable, irrespective of the expectations scheme assumed. For the annual data, the results on the full sample are mixed, accepting consistency with GARP under one expectations scheme but rejecting it under the other.[12] For the most part, then, the data seem inconsistent with a representative agent maximizing an unchanging, intertemporally separable utility function over the full sample periods. The data, however, could be consistent with this type of representative-agent model if the parameters or the functional form of the utility function changed appropriately at the breakpoints indicated earlier. Interestingly, the annual and quarterly data with perfect foresight nearly agree on whether and when a break occurs. Under static expectations, the annual and quarterly samples are continuous over 1960–1990, whereas under perfect foresight they both show a break in 1981. The 1981 break is especially strong in the quarterly data, which show a discontinuity between the 1980.4 and 1981.4 observations and which violate GARP for every possible pairing within the discontinuity period. The strength of the quarterly results and their consistency with the annual results suggests something more than just an artifact of measurement error. Also, this pattern of rejections is consistent with preferences being constant over the four quarters of the year for years on either side of the 1980–1981 period. Previous studies [e.g., Browning (1989)], hypothesized that GARP would fail in quarterly data because preferences might be periodic over the quarters, but the data do not support that hypothesis. The monthly data do not show breaks at the same points as the annual and quarterly data, but there is some reason to believe that the monthly data simply are inadequate, as has been noted in the literature (Wilcox, 1992) and as we discuss in more detail later.

Second, we never find a break in 1979 for the annual or quarterly data. The supposed change in Federal Reserve operating procedure that year often is cited as a possible source of instability in consumption-based asset pricing models [e.g., Ghysels and Hall (1990)]. Our results suggest that there is no problem with 1979 for consumption analysis, although there are problems with other years.

Third, the results suggest special problems with the monthly data. Under static expectations, there is no GARP-consistent subsample long enough to use in an econometric study. The results for perfect foresight are much less drastic, but even there the longest period is only 24 years. For "short-run" kinds of questions, 24 years (or 288 months) may be a sufficiently long sample, but it is unlikely to be

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

adequate for "long-run" issues where the span of the data is more important than the frequency of observation. We discuss the monthly data further in the next section.

Finally, there is no systematic relation between the expectation schemes and the test results. For monthly data, perfect foresight gives fewer rejections than static expectations, but the opposite is true for the annual and quarterly data.

4.2. Separability Tests

The separability tests are an extension of the GARP tests (see Varian's theorem, above) and so can be done only on data consistent with GARP. We therefore restrict attention to the sample periods listed in Table 2.

We examine the separability assumptions most frequently used in the consumption literature. Our annual data set reports expenditure on 62 categories of consumption goods—17 types of nondurables, 34 types of services, and 11 types of durables; the quarterly and monthly data have fewer series (see Appendix B, Section B.3). In conducting our tests, we construct subsets (groups) of these categories but at no time do we aggregate them, thus avoiding any possible problems associated with a particular method of aggregation.[13] We perform our tests under both static expectations and perfect foresight. The results are reported in Table 3.

The annual and quarterly data tell the same story, without much difference between the two expectations schemes. The first hypothesis, S1, is that the three major categories of consumption expenditure are mutually separable from each other. That hypothesis does not pass either the necessary or the necessary and sufficient conditions, implying that at least one of the categories of expenditure cannot be analyzed independently of the other two.

The next three hypotheses examine the possibility that durables are separable from nondurables and services. The upshot is that they are not. Hypothesis S2 proposes mutual separability of nondurables and services as a group on the one hand and durables on the other, S3 is the same as S2 except that clothing and shoes are reclassified as durables rather than nondurables as suggested by Darby (1974), and S4 weakens S2 to one-sided separability of durables from the group of nondurables and services. None of the three passes the necessary and sufficient conditions; only S3 passes the necessary conditions and then only for annual data under perfect foresight.

The remaining hypotheses examine if nondurables alone or nondurables and services together are separable from durables. The first three of these hypotheses—S5, S6, and S7—give inconclusive results, usually passing the necessary conditions but failing the necessary and sufficient conditions. These results looked promising, and so, we explored further by ungrouping the durables category into its components. Space prohibits reporting all the combinations that we tried, and so, we report the successes and some close relatives. Note that in Table 3 the first seven separability tests (S1–S7) are the same for all frequencies of data. For the annual data, hypothesis S8 passed the necessary and sufficient conditions for both expectations schemes. That hypothesis splits durables into motor vehicles, other autos, and all

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

Table 3. Tests for weak separability

Separability hypothesis	Expectations model	
	Static expectations	Perfect foresight
Annual data		
S1: $f[u_1(\text{nond}), u_2(\text{ser}), u_3(\text{dur})]$	X	X
S2: $f[u_1(\text{nond, ser}), u_2(\text{dur})]$	X	X
S3: $f[u_1(\text{nond-clothing, ser}), u_2(\text{dur, clothing})]$	X	N
S4: $f[\text{nond, ser}, u_1(\text{dur})]$	X	X
S5: $f[u_1(\text{nond, ser}), \text{dur}]$	N&S	N
S6: $f[u_1(\text{nond}), u_2(\text{ser, dur})]$	N&S	X
S7: $f[u_1(\text{nond}), \text{ser, dur}]$	N	N
S8: $f[u_1(\text{nond, ser}), u_2(\text{motor vehicles}), u_3(\text{other autos}), u_4(\text{remaining nine durables})]$	N&S	N&S
S9: $f[u_1(\text{nond, ser}), u_2(\text{motor vehicles, other autos}), u_3(\text{remaining nine durables})]$	N	X
S10: $f[u_1(\text{nond}), u_2(\text{ser}), u_3(\text{motor vehicles}), u_4(\text{other autos}), u_5(\text{remaining nine durables})]$	X	X
Quarterly data		
S1: $f[u_1(\text{nond}), u_2(\text{ser}), u_3(\text{dur})]$	X	X
S2: $f[u_1(\text{nond, ser}), u_2(\text{dur})]$	X	X
S3: $f[u_1(\text{nond-clothing, ser}), u_2(\text{dur, clothing})]$	X	X
S4: $f[\text{nond, ser}, u_1(\text{dur})]$	X	X
S5: $f[u_1(\text{nond, ser}), \text{dur}]$	N	N
S6: $f[u_1(\text{nond}), u_2(\text{ser, dur})]$	N	N
S7: $f[u_1(\text{nond}), \text{ser, dur}]$	N	N
S11: $f[u_1(\text{nond}), u_2(\text{ser}), u_3(\text{motor vehicles}), u_4(\text{furniture, other durables})]$	X	X
S12: $f[u_1(\text{nond, ser}), u_2(\text{motor vehicles, furniture}), u_3(\text{other dur})]$	X	X
S13: $f[u_1(\text{nond, ser}), u_2(\text{motor vehicles}), u_3(\text{furniture, other dur})]$	N&S	N&S
Monthly data		
S1: $f[u_1(\text{nond}), u_2(\text{ser}), u_3(\text{dur})]$	X	X
S2: $f[u_1(\text{nond, ser}), u_2(\text{dur})]$	X	X
S3: $f[u_1(\text{nond-clothing, ser}), u_2(\text{dur, clothing})]$	X	X
S4: $f[\text{nond, ser}, u_1(\text{dur})]$	X	X
S5: $f[u_1(\text{nond, ser}), \text{dur}]$	X	X
S6: $f[u_1(\text{nond}), u_2(\text{ser, dur})]$	X	X
S7: $f[u_1(\text{nond}), \text{ser, dur}]$	X	X
S11: $f[u_1(\text{nond}), u_2(\text{ser}), u_3(\text{motor vehicles}), u_4(\text{furniture, other durables})]$	X	X

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

TABLE 3. (Continued.)

Separability hypothesis	Expectations model	
	Static expectations	Perfect foresight
S12: $f[u_1$(nond, ser), u_2(motor vehicles, furniture), u_3(other dur)]	X	X
S13: $f[u_1$(nond, ser), u_2(motor vehicles), u_3(furniture, other dur)]	N*	X

NOTES: Annual data—dur, nond, ser = {11 durable goods}, {17 nondurable goods}, {34 services}; N = passes necessary condition; N&S = passes necessary condition and a sufficient condition; X = fails all conditions. Test results apply to all subsample periods except when marked by an asterisk (*), in which case the separability test failed for the sample period 1990.07–1990.12.

remaining durables. These three subcategories and the group of nondurables and services are all mutually separable. Hypothesis S9 is but one example showing that even a slight attempt to consolidate these categories leads to much weaker test results. Hypothesis S10 tries to separate nondurables and services, given the durables divisions of S8, but is rejected. For the quarterly data, hypotheses S11 and S12 both fail all tests, but the closely related hypothesis S13 passes the necessary and sufficient conditions for both expectations schemes. Hypothesis S12 shows that a slight reformulation of S13 leads to rejection.

The main implication of these results is that, for annual or quarterly data, an econometric analysis of nondurables and services as a group, paying no attention to durables, is valid but analysis of either nondurables alone or durables alone is not. Also, when analyzing all three categories at once, one must divide durables into the subcategories shown in hypotheses S8 and S13. It is of interest to compare our results with Wilcox's (1992) recommendations based on an analysis of the measurement system for consumption data. His analysis suggested that one may wish to disaggregate consumption expenditure into (at least) three groups: motor vehicles, services, and the rest. Our results provide empirical support for the treatment of motor vehicles as a separate group, but reject the treatment of services as a separate group.

Finally, let us return to the results for the monthly data. None of the separability hypotheses pass the necessary and sufficient conditions; only one hypothesis passes the necessary conditions, and even then not over all sample periods. These results are in striking contrast to those for the annual and quarterly data, and there are at least two ways to interpret them. The first interpretation accepts the results at face value, in which case analysis of monthly data cannot ignore any category of consumption expenditure in examining the behavior of any other category. The second interpretation is that the results are artifacts of bad data. Recall that it is difficult to find subsamples of the monthly data that pass GARP. Perhaps those difficulties and the monthly data's different separability implications both reflect the problems reported by Wilcox (1992) with the methods used to construct the

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

monthly data. If so, the monthly data may be of questionable value for analyzing consumption based on the representative-agent model.

5. CONCLUDING REMARKS

We have used Varian's (1982, 1983) nonparametric tests to examine two issues pertaining to the representative-agent model. First, we use the GARP tests to investigate whether aggregate consumption data for the United States could have been generated by a representative agent whose utility function is intertemporally separable. Second, we explore the separability structure of the representative agent's utility function. Although the nonparametric tests have some limitations, the alternative of not testing at all obviously also has some shortcomings.

The GARP tests almost uniformly find that the data are inconsistent with a representative-agent model with intertemporally separable utility over the full period of available data. The exception is annual data under static expectations. Annual data under perfect foresight and quarterly data under either scheme show a break in behavior around 1981, which suggest that the representative-agent model with intertemporally separable utility either is wrong altogether or at least that the utility function changed after 1981. Finally, the monthly data exhibit so many violations of GARP that they cannot be consistent with an intertemporally additive utility function that is constant for any appreciable length of time.

Many empirical studies of consumption based on this type of representative-agent model use samples over which our tests reject GARP. One interpretation of our results is that those studies use misspecified models. However, such a conclusion may be premature. Because a rejection with the GARP test is all or nothing, we cannot assess whether the violations are more likely to be due to measurement error or some more fundamental change in the model, nor can we assess the likely impact of violations of GARP on the estimation of the parameters of Euler equations derived from the representative-agent model. All of these issues are important areas for future research. Until such work becomes available, however, we believe that our results may be best regarded as a nonparametric structural stability test that suggests caution in estimating and interpreting any representative-agent model estimated with annual or quarterly data that include observations for 1981 or estimated with any set of monthly data at all.

An alternative explanation for our results is that the representative agent has preferences that vary seasonally (thus causing the large number of GARP rejections in the monthly data) or has a utility function that is not intertemporally separable. An associate editor has pointed out that testing for seasonality in the monthly data could be done by creating twelve data sets, one comprising all of the January observations, another comprising all of the February observations, and so on, and then testing each of the twelve data sets for consistency with GARP. GARP consistency of the twelve separate data sets would be strong evidence that the rejections in the total data set were the result of seasonally varying preferences. This extensive testing would be an interesting topic for future research. With respect

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

to intertemporal separability, it is entirely possible that the data are consistent with a structurally constant utility function possessing such features as habit persistence. That is a plausible explanation for the high number of violations of GARP with the monthly data, although there are other grounds for suspecting that the monthly data are just of poor quality (Wilcox, 1992). However, it then is very striking that we found so few violations of GARP with the quarterly and annual data, suggesting that any habit formation effects persist for less than three months.

Let us turn now to the separability tests. These tests can be applied only to data sets that pass GARP, and so, for the rest of this discussion we focus exclusively on those subsets of the annual and quarterly data consistent with GARP. The results show that the main categories of consumption (nondurables, services, and durables) are not mutually separable, and so, one should not analyze arbitrary subsets of the consumption data. However, nondurables and services are separable from durables, although not vice versa. This result implies that nondurables and services can be analyzed together as a group independently of durables, but not vice versa. Also, we find evidence that motor vehicles should be treated as a separate group.

Most studies use a single aggregate of nondurable and services expenditures as their measure of consumption and ignore durables. According to our separability results, that procedure is legitimate for annual and quarterly data. If one wants to include durables, however, our separability tests reject forming a single aggregate of durable goods. Rather, one should divide durables into some combination of motor vehicles, other autos, furniture, and other durables, the exact combination depending on whether the data are quarterly or annual.

NOTES

1. See Swofford and Whitney (1987, 1988), Manser and McDonald (1988), and Browning (1989) for pioneering work in applying GARP tests to aggregate consumption data. These studies have some limitations, which we discuss later. Belongia and Chalfant (1989) use the GARP tests to study the empirical definition of money, and Chalfant and Alston (1988) have used them to study changes in tastes for meat.

2. The numbers U and λ can be interpreted as the utility levels and marginal utilities of income at the observed demands. The inequality in part 1 follows directly from the concavity of the utility function and the first-order conditions of utility maximization. See Varian (1982, 1983).

3. Note that there are two kinds of aggregation going on. Because we are talking about a representative consumer, the elements of x—that is, the individual consumption goods—already have been aggregated across households. We now are considering the possibility of aggregating across the goods themselves. The best way to do the latter is a matter of some dispute, and several methods have been proposed. It is not our purpose to study aggregation methods, and so, we simply assume that some method has been chosen. Later, we remark on the implications of our empirical results for making that choice.

4. See Blackorby et al. (1978) for more discussion of separability.

5. Use of this simplifying procedure is routine in both Euler equation and consumption function studies. For example, Hall (1978) uses nondurables plus services and Flavin (1981) uses nondurables alone.

6. Swofford and Whitney (1994) develop a test of the necessary and sufficient conditions in Varian's theorem. Unfortunately, it is computationally very intensive and cannot be performed for samples of the sizes in our paper; see Swofford and Whitney (1994) for further discussion.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

7. Varian (1985) and Epstein and Yatchew (1985) have extended Varian's (1982, 1983) test to a stochastic environment by allowing for the presence of measurement error. Their implementation requires certain assumptions about the measurement-error process, including knowledge of its variance, and so, we do not pursue those tests here. Bronars (1987) reports simulation evidence on the rejection frequencies of Varian's GARP test when the consumption data are random. His design assumes nine categories of consumption goods and 31 years of annual data. His results indicate that the power is in excess of 0.91 against this alternative when per capita consumption data are used.

8. We ignore the possibility that both nondurables and services as measured by the National Income Accounts contain durable elements.

9. The use of user costs is one of the features that distinguishes our work from that of Manser and McDonald (1988), who use the price of new durables rather than the user cost.

10. For example, Diewert (1974) and Swofford and Whitney (1987, 1988) set δ at 0.10 for all goods, and Christensen and Jorgenson (1969) set it at 0.20.

11. Negative user costs have the implausible implication that the good is more than free: its use pays rather than costs the user.

12. This conclusion differs from that of Manser and McDonald (1988), who fail to reject GARP for annual data (the only data they test). Our tests show that whether or not the annual data are consistent with GARP depends on the expectations scheme, which affects the user cost of durables. Manser and McDonald use new durables' prices rather than user costs as the price of durables and so never address the expectations problem. However, our results agree with those of Browning (1989), who reported that the annual data for 1954–1979 satisfy GARP.

13. In contrast, Swofford and Whitney (1987, 1988) use only the aggregate totals for nondurables, services, and durables reported in the NIPA data. Those aggregates presuppose an amount of separability not consistent with the data, as we show later. They also are created by the simple-sum method, which seems inappropriate in light of our separability results.

REFERENCES

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

Barnett, W. & S. Choi (1989) A Monte Carlo study of tests of blockwise weak separability. *Journal of Business and Economics Statistics* 7, 363–377.

Belongia, M.T. & J.A. Chalfant (1989) The changing empirical definition of money: Some estimates from a model of the demand for money substitutes. *Journal of Political Economy* 97, 387–397.

Blackorby, C., D. Primont & R.R. Russel (1978) *Duality, Separability and Functional Structure: Theory and Economic Applications*. Amsterdam: North-Holland.

Bronars, S.G. (1987) The power of nonparametric tests of preference maximization. *Econometrica* 55, 693–698.

Browning, M. (1989) A nonparametric test of the life-cycle rational expectations hypothesis. *International Economic Review* 30, 979–992.

Campbell, J. & N.G. Mankiw (1989) Consumption, income, and interest rates: Reinterpreting the time series evidence. In O.J. Blanchard & S. Fischer (eds.), *NBER Macroeconomics Annual 1989*. Cambridge, MA: MIT Press.

Chalfant, J.A. & J.M. Alston (1988) Accounting for changes in tastes. *Journal of Political Economy* 96, 391–410.

Christensen, L. & D.W. Jorgenson (1969) The measurement of U.S. real capital input. *Review of Income and Wealth* 15, 293–320.

Darby, M.R. (1974) The permanent income theory of consumption—A restatement. *Quarterly Journal of Economics* 88, 228–250.

Deaton, A. & J. Muellbauer (1980) *Economics and Consumer Behavior*. Cambridge, UK: Cambridge University Press.

Diewert, W.E. (1973) Afriat and revealed preference theory. *Review of Economic Studies* 40, 419–426.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

Diewert, W.E. (1974) Intertemporal consumer theory and the demand for durables. *Econometrica* 42, 497–516.
Epstein, L.G. & A.J. Yatchew (1985) Non-parametric hypothesis testing procedures and applications to demand analysis. *Journal of Econometrics* 30, 149–169.
Flavin, M.A. (1981) The adjustment of consumption to changing expectations about future income. *Journal of Political Economy* 89, 974–1009.
Ghysels, E. & A.R. Hall (1990) Are consumption based asset pricing models structural? *Journal of Econometrics* 46, 121–139.
Gorman, W.M. (1959) Separable utility and aggregation. *Econometrica* 27, 469–481.
Hall, R.E. (1978) Stochastic implications of the life-cycle permanent income hypothesis: Theory and evidence. *Journal of Political Economy* 86, 971–987.
Hicks, J.R. (1956) *Value and Capital*. Oxford: Oxford University Press.
Kirman, A.P. (1992) Whom or what does the representative individual represent? *Journal of Economic Perspectives* 6, 117–136.
Kocherlakota, N. (1996) The equity premium puzzle: It's still a puzzle. *Journal of Economic Literature* 34, 42–71.
Manser, M.E. & R.J. McDonald (1988) An analysis of substitution bias in measuring inflation, 1959–85. *Econometrica* 56, 909–930.
McCulloch, H.J. & H.-C. Kwon (1993) U.S. Term Structure Data, 1947–1991. Working paper 93-6, Ohio State University.
Swofford, J.L. & G.A. Whitney (1987) Non-parametric tests of utility maximization and weak separability for consumption, leisure, and money. *Review of Economics and Statistics* 69, 458–464.
Swofford, J.L. & G.A. Whitney (1988) A comparison of non-parametric tests of weak separability for annual and quarterly data on consumption, leisure, and money. *Journal of Business and Economics Statistics* 6, 241–246.
Swofford, J.L. & G.A. Whitney (1994) A revealed preference test for weakly separable utility maximization with incomplete adjustment. *Journal of Econometrics* 60, 235–249.
Thornton, D. & P. Yue (1992) An extended series of Divisia monetary aggregates. *Quarterly Review of Federal Reserve Bank of St. Louis* 74, 35–52.
Varian, H. (1982) The nonparametric approach to demand analysis. *Econometrica* 50, 945–973.
Varian, H. (1983) Nonparametric tests of consumer behavior. *Review of Economic Studies* 50, 99–110.
Varian, H. (1985) Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics* 30, 445–458.
Wilcox, D.W. (1992) The construction of U.S. consumption data: Some facts and their implications for empirical work. *American Economic Review* 82, 922–941.

APPENDIX A: SARP AND GARP

A.1. DEFINITIONS

1. x_i is directly revealed as preferred to x, written $x_i R^0 x$, if $p_i x_i \geq p_i x$.
2. x_i is strictly directly revealed as preferred to x, written $x_i P^0 x$, if $p_i x_i > p_i x$.
3. x_i is revealed as preferred to x, written $x_i R x$, if $p_i x_i \geq p_i x_j$, $p_j x_j \geq p_j x_k, \ldots, p_m x_m \geq p_m x$ for some sequence of observations $\{x_i, x_j, \ldots, x_m\}$.
4. x_i is strictly revealed as preferred to x, written $x_i P x$, if there exist observations x_j and x_k such that $x_i R x_j$, $x_j P x_k$, $x_k R x$.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

A.2. AXIOMS

1. SARP. The following statements of SARP are equivalent.
   (A) $x_i R x_j$ and $x_j R x_i \Rightarrow x_i = x_j$.
   (B) $x_i R x_j$ and $x_i \neq x_j \Rightarrow$ not $x_j R x_i$.
   (C) $x_i R x_j$ and $x_i \neq x_j \Rightarrow$ not $x_j R^0 x_i$.
2. GARP.
   $x_i R x_j \Rightarrow$ not $x_j P^0 x_i$.

APPENDIX B

B.1. ANNUAL DATA

Population. All stocks and flows are measured in per-capita terms. Population is measured as Total Population from the *Report of the Council of Economic Advisors*, various issues.

Consumption expenditure. For 1929–1982, quantities and prices are from the *National Income and Product Accounts*, U.S. Personal Consumption Expenditures, Tables 2.6 and 2.7. For 1983–1990, the data are from the Bureau of Economic Analysis tapes. Quantities are real 1987 dollars, and prices are normalized to a base of 1987 = 100. The earlier data are reported in 66 categories of consumption expenditure. The later data are divided into more detailed categories than that, and so we aggregated them up to the earlier categories. Our analysis uses all 17 categories of nondurable expenditure and 34 of the 36 categories of service expenditure, omitting net foreign remittances and net foreign travel, which sometimes are negative and thus unusable. The latter two series together always amount to less than 1% of total consumption expenditure.

Although the NIPA data report 13 types of durable-goods expenditure, the *Survey of Current Business* reports only 11 stock series. We therefore aggregated up the relevant categories in the NIPA data to match those of the SCB.

Depreciation rates. The following service lives, in years, are from the *Fixed and Reproducible Tangible Wealth*: autos (10), other motor vehicles (8), furniture (14), kitchen and other household appliances (11), china (10), other durable households (10), radio and musical instruments (9), jewelry (11), ophthalmic products and orthopedic appliances (6), books (10), and wheel goods (10). We use the reciprocal of a durable's expected number of service years as the depreciation rate for that good.

Interest rate. The interest rate is the six-month prime commercial paper rate, taken from the *Survey of Current Business*. We use this rate because it is the closest series to a one-year rate dating back to 1929.

B.2. QUARTERLY DATA

Consumption expenditure. Data are from the *National Income and Product Accounts, Private* Consumption Expenditure, Seasonally Adjusted Quarterly Totals at Annual Rates, 1959.1–1990.4 Constant Cost (billions of 1987 dollars). The quarterly NIPA data have only 14 disaggregated series.

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press

We calculate the quarterly stocks of consumer durables from the annual data as in Campbell and Mankiw (1989) because there are no quarterly data available. We measure the quarterly stock of durable goods as the average of the end-of-quarter stock and previous end-of-quarter stock. We use the annual stock for 1958 as the end-of-quarter stock for the initial quarter 1959.1. We then construct the end-of-quarter stocks in the usual way by depreciating the previous quarter's stock and adding the current quarter's expenditure on the durable in question. We calculate the depreciation rate δ, from $k_t = d_t - (1 - \delta)k_{t-1}$, by selecting a depreciation rate for each category until the stock in 1990.4 is close to the annual stock for 1990. The depreciation rates (in percent) per quarter are total durables (6) [the same rate as that of Campbell and Mankiw (1990)], motor vehicles and parts (5.9), furniture and household equipment (4.2), and other durables (4.6). These depreciation rates may seem high, but note that they ignore discards. Annual stock measures are net of discards, and so, their depreciation rates are necessarily lower than those for the gross quarterly stocks.

Interest rate. The interest rate is from McCulloch and Kwon (1993). This monthly interest rate has three months to maturity and is aggregated over (Dec, Jan, Feb), (Mar, Apr, May), (June, July, Aug), (Sep, Oct, Nov) into a quarterly series. (This interest rate is not used for the annual data because it does not date back to 1929.)

B.3. MONTHLY DATA

The monthly data set is created using the same technique as in the quarterly data set. The PCE monthly series are from Citibank Database for 1959:01–1990:12. The following series are used:

GMCDAQ—motor vehicles and parts
GMCDFQ—furniture and household equipment
GMCDOQ—other durables
GMCNCQ—clothes and shoes
GMCNFQ—food
GMCNGQ—gasoline and oil
GMCNOQ—other nondurable goods
GMCNOFQ—fuel, oil, and coal
GMCSHQ—housing services
GMCSGQ—housing operations
GMCSMQ—medical care services
GMCSOQ—other services
GMCSTQ—transportation services
P16—total civilian population
INT—interest rate is from Thornton and Yue (1992) and is the highest rate that the agent can obtain in financial markets

https://doi.org/10.1017/S1365100500016035 Published online by Cambridge University Press