

GOODNESS-OF-FIT IN OPTIMIZING MODELS

Hal R. VARIAN*

University of Michigan, Ann Arbor, MI 48109, USA

Conventional econometric tests of optimizing models typically involve embedding the optimizing model in a parametric specification and then examining the parametric restrictions imposed by the optimization hypothesis. The optimization hypothesis is rejected if the estimated parameters are significantly different, in the statistical sense, from the values implied by optimization. I argue that a more fruitful approach to testing optimizing behavior is to measure the departure from optimization using the estimated objective function, and see whether this departure is significant in an *economic* sense. I discuss procedures for doing this that can be used in several sorts of optimizing models, and give a detailed illustration in the case of aggregate demand estimation.

1. Introduction

Much of economics rests on the principle of optimizing behavior. Firms are assumed to minimize costs and maximize profits; consumers are assumed to maximize utility; and so on. In the last several years, standard techniques have been developed to test these models of optimizing behavior. Suppose, for example, that we are attempting to test the hypothesis that a time series of observations on factor choices by a firm can be viewed as cost-minimizing behavior. A common approach would be to pick a parametric form for the underlying cost or production function, derive the associated set of factor demand functions, and then see if the estimated parameters satisfy the restrictions imposed by the model of cost minimization.

Similarly, if one wanted to test a set of data on consumer choices for consistency with utility maximization behavior, one would first specify a function form for the utility function, derive the associated set of utility-maximizing demand functions, estimate the parameters of these demand functions using the consumer choice data, and then see if these estimated parameters satisfy the restrictions imposed by the model of utility maximization.

In my view, these procedures are not very good ways to test models of optimizing behavior for two distinct reasons. First, there is often no need to

*I would like to thank Wei Li for programming assistance and Eduardo Ley for proofreading. This work was supported in part by the National Science Foundation. This paper was prepared for the conference on Parametric and Nonparametric Approaches to Frontier Analysis, held at the University of North Carolina, September 29–October 1, 1988.

embed the optimizing model in a parametric framework. I argue below that it is perfectly possible to test reasonably complex models of optimization behavior without having to use parametric specifications. Second, testing parametric restrictions by using classical significance tests involve an overly restrictive sense of ‘significance’. What matters for most purposes in economics is not whether a consumer’s violation of the optimizing model is *statistically* significant, but whether it is *economically* significant. And the economic significance of a departure from optimizing behavior has nothing to do with whether or not estimated parameters pass or fail a test of statistical significance.

Hence the conventional methods are lacking in two senses: first, they have an excess reliance on parametric forms, and second, they test for statistically significant violations of optimization rather than economically significant violations. Let us examine each of these points in more detail.

2. Nonparametric tests of optimizing behavior

Suppose that we observe a set of price vectors, p^t , and net output vectors, y^t , for $t = 1, \dots, T$, and want to test the hypothesis of period-by-period profit maximization. Then a necessary condition for these data to be consistent with profit maximization is that the following inequalities are satisfied:

$$p^t y^t \geq p^s y^s,$$

for all pairs of observations s and t . These inequalities simply say that the profit from the observed choices must be at least as large as the profits from any other feasible choice. Varian (1984) refers to this as the Weak Axiom of Profit Maximization (WAPM). Similar inequalities have been examined by several other authors including Afriat (1972) and Samuelson (1947).

It can also be shown that WAPM is a *sufficient* condition for profit maximization in the sense that any set of data that satisfies WAPM can be used to construct a ‘nice’ production set that could have generated the observed behavior as optimizing behavior. See Varian (1984) for details.

Hence a sensible test of optimizing behavior in this context is simply to see if the observed prices and net output vectors satisfy the inequalities implied by WAPM. If the data violate the inequalities, then we reject the model of optimizing behavior.

Suppose that we observe a set of data (w^t, x^t, y^t) for $t = 1, \dots, T$, where w^t is a vector of factor prices, x^t is a vector of factor demands, and y^t is a (scalar) measure of output. We might be interested in testing the hypothesis that the firm that generated this data is minimizing the cost of producing the observed output.

If the firm is minimizing costs, it must satisfy the following set of inequalities:

$$w^t x^t \leq w^t x^s \quad \text{for all } y^s \geq y^t.$$

These inequalities require that the cost of the observed production plan must be no greater than the cost of any other production plan that produces at least as much output. Varian (1984) calls this the Weak Axiom of Cost Minimization (WACM).¹

Again, this condition is necessary and sufficient for cost-minimizing behavior in the sense that, if some data satisfy WACM, then it is possible to construct a production set that would generate the observed choices as cost-minimizing choices. It is very easy to apply this test to observed choices to see if they violate the inequalities; no appeal to parametric methods is required.

Finally, suppose that we observe some price vectors p^t and quantity vectors x^t , for $t = 1, \dots, T$, and want to test the hypothesis that these data were generated by a utility-maximizing consumer. Define the revealed preference relation R by $x^t R x^s$ if and only if there is some sequence of observations $x^r \dots x^u$ such that $p^t x^t \geq p^r x^r, \dots, p^u x^u \geq p^s x^s$. Then a set of data is consistent with the model of utility maximization if and only if it satisfies the Generalized Axiom of Revealed Preference (GARP),

$$x^t R x^s \quad \text{implies} \quad p^s x^s \leq p^s x^t.$$

Again, this condition is easily tested; see Varian (1982a) for details.

Given that each of these classical models of optimizing behavior is easily tested by simply checking a set of inequalities, why do the conventional procedures use complicated statistical measures? Certainly the inequalities described above have been in the literature a long time. See Samuelson (1938), Afriat (1967), Diewert and Parkan (1985), and others.²

One explanation is that economists are simply not used to thinking about the implications of optimizing models for a finite set of observations. It is more natural for economists, perhaps, to think of the outcome of optimization to be an entire demand or supply *function*.

Another, perhaps more important explanation, is that the nonparametric tests described above are 'sharp' tests: either the data pass the test *exactly*, or

¹See Varian (1984) for a discussion of the literature on this sort of test, which includes contributions from Samuelson (1947), Hanoeh and Rothschild (1972), and Afriat (1972).

²For more recent work in nonparametric analysis of consumption behavior, see Browning (1984), Bronars (1987), Deaton (1985), Green and Srivastava (1985, 1986), Houtman and Maks (1987), Landsburg (1981), Manser and McDonald (1988), and Varian (1982a, 1982b, 1984, 1985, 1988).

they don't. If the data don't satisfy the test, the optimizing model is rejected: the tests do not allow for an 'error term'.

There have been some attempts to deal with this problem of the overly sharp nature of nonparametric tests. Banker and Maindiratta (1988) suggest finding the largest set of observations consistent with optimization. Varian (1985) suggests finding the set of data that is nearest to the observed data in some appropriate norm.

Conventional statistical tests *do* allow for an error term. To test the hypothesis that some relationship holds among some estimated parameters, we ask whether the value of some test statistic is likely or unlikely according to the sampling distribution of the parameters. Roughly speaking, the optimization model is rejected if the observed value of the test statistic is unlikely.

The problem with this procedure, in my view, is that it has little to do with the *economic* significance of the violation. For example, optimization of some particular parametric form may imply that two parameters should sum to one. If we test this hypothesis and reject it, we must reject the optimizing model. But what are we rejecting? Exact optimization implies that the two parameters must sum to *exactly* one. But *exact* optimization isn't a very interesting hypothesis. It is very unlikely that firms *exactly* maximize profits or minimize costs; it is even more unlikely that consumers *exactly* maximize utility. It is especially unlikely that consumers maximize some arbitrary parametric approximation to utility.

What we usually care about is whether optimization is a *reasonable* way to describe some behavior. For most purposes, 'nearly optimizing behavior' is just as good as 'optimizing' behavior.³

The conventional parametric tests miss this distinction: given enough data, we can always reject nonoptimizing behavior, even if it is 'nearly optimizing behavior'. The value of the test statistic will typically give no clue as to whether the economic agent under examination is nearly optimizing or grossly nonoptimizing.⁴

3. Goodness-of-fit measures

An alternative approach to testing optimizing behavior is to ask how large the violations of relevant inequalities are in terms of a reasonable *economic*

³See Akerlof and Yellen (1985) and Cochrane (1989) for interesting discussions of nearly optimizing behavior.

⁴A similar point made by McCloskey (1985, 1989) in a somewhat different context. McCloskey points out that significance testing, as commonly used in economics, does not provide an appropriate measure of the 'importance' of a variable in a regression. But McCloskey's critique applies more broadly; most statistical tests measure violations of a hypothesis in terms of the sampling distribution of the test statistic, and this is rarely a useful measure of the *importance* of this violation.

norm. For example, suppose that we observe some violations of the Strong Axiom of Profit Maximization. That is, we observe a pair of observations s and t for which

$$p^t y^t < p^t y^s.$$

This inequality says that the firm could make more profit by choosing y^s when in fact it chose y^t . In this case, a reasonable measure of the magnitude of the violation of profit-maximizing behavior is

$$r^{ts} = p^t (y^s - y^t) / p^t y^t = p^t y^s / p^t y^t - 1.$$

This is simply the percent *extra* profit that the firm could have made at the prices p^t if it had chosen the production vector y^s , rather than the production vector y^t .

The numbers r^{ts} should be interpreted as ‘residuals’ appropriate for examining the optimization model. The best way to present these residuals might be to list the observations and indicate next to each one the magnitude of the foregone profit. Or, one might want to look at the average value of the foregone profit, or the largest value of the foregone profit. Any of these numbers would be a reasonable way to measure how ‘close’ the observed behavior comes to profit-maximizing behavior.

The case of cost maximization is almost the same. If we have a violation of WACM, we have two observations t and s such that

$$w^t x^t > w^t x^s \quad \text{for some } y^s \geq y^t.$$

In this case,

$$w^t (x^t - x^s) / w^t x^t = 1 - w^t x^s / w^t x^t$$

is a reasonable measure of the departure from cost minimization. This number simply measures how much the firm could have saved if it had chosen x^s rather than x^t when it faced factor prices w^t . Again, one might choose the average value or the maximum value of this index as a measure of the degree of violation of maximization.

If these numbers are small, then it seems reasonable to think of the firm under consideration as being ‘more-or-less’ an optimizing firm.⁵ True, it isn’t exactly optimizing, but *exactly* optimizing behavior isn’t a very plausible hypothesis to begin with.

In addition, the distribution of these measures of profit maximization or cost minimization may be of considerable interest themselves. Suppose, for

⁵How small is small? In general this depends on the problem at hand. The ‘magic number’ of significance tests, 5%, is probably a reasonable choice.

example, that we are examining the case of profit maximization using data on a single firm, and we find that most violations of WAPM indicate that the firm would be better off at time t making a choice that was made at some later date. This suggests that technological progress or learning-by-doing may be involved: the more profitable choices weren't made at time t because they weren't feasible.

Or suppose that we are examining cross-sectional data and we find that most of the violations of WAPM involve a single firm. This might be taken as evidence that this firm really doesn't have access to the same technology as the others. The *pattern* of violations can tell us a lot about what is going on in the data.

4. Consumer choice

The description of a reasonable measure of goodness-of-fit in the case of consumer choice is somewhat more involved. We follow the suggestion of Afriat (1967).

Afriat's measure is calculated in the following manner. For a given set of numbers (e^t), $t = 1, \dots, T$, with $0 \leq e^t \leq 1$, define an extension of the standard direct revealed preference relation by

$$x^t R_e^0 x^s \quad \text{if and only if} \quad e^t p^t x^t \geq p^t x^s.$$

If $e^t = 1$, this is the standard direct revealed preference relation; if $e^t = 0$, the relation is vacuous in the sense that observation t cannot be revealed preferred to any other observation. As e^t varies from 1 to 0 the number of observations revealed preferred to other observations monotonically decreases.

We refer to e^t as the *Afriat efficiency index* for observation t . It can be thought of as how much less the potential expenditure on a bundle x^s has to be before we will consider it worse than the observed choice x^t . If e^t is 0.90, for example, we will only count bundles whose cost is less than 90% of an observed choice as being revealed worse than that choice. Said another way: if e^t is 0.90 and x^s would cost only 5% less than x^t , we would not consider this a significant enough difference to conclude that x^t was preferred by the consumer to x^s . We are allowing the consumer a 'margin of error' of $(1 - e^t)$.

Given an arbitrary set of data (p^t, x^t) , let us choose a set of efficiency indices (e^t) that are as close as possible to 1 in some norm. If the data satisfy the revealed preference conditions exactly, then we can choose $e^t = 1$ for all $t = 1, \dots, T$. If we choose $e^t = 0$ for all $t = 1, \dots, T$, then the data vacuously satisfy the revealed preference conditions, since no observation is revealed preferred to any other. Thus for any reasonable norm, there will be some set

of (e^t) that are as close as possible to 1 that will summarize 'how close' the observed choices are to maximizing choices.

In Afriat's (1967) original treatment of this idea, he considered choosing a single e that applied to all observations, rather than a different e^t for each observation. The advantage of Afriat's original proposal is that it is much easier to compute a single index e than the multiple indices (e^t) .

Houtman and Maks (1987) suggest the following binary search. Start with $e = 1$ and test for violations of revealed preference using Warshall's algorithm as described in Varian (1982a). If the data fail to satisfy the strong axiom, try $e = \frac{1}{2}$. If $e = \frac{1}{2}$ doesn't work, try $e = \frac{1}{4}$. If $e = \frac{1}{2}$ does work, try $e = \frac{3}{4}$, and so on. After n revealed preference tests, you are within $1/2^n$ of the actual efficiency index.

Computing the set of efficiency indices that are as close as possible to 1 in some norm is substantially more difficult. If we choose a quadratic norm, for example, we would have to solve a problem such as

$$E = \min_{(e^t)} \sum_{t=1}^T (e^t - 1)^2, \quad (1)$$

subject to the constraint that the revealed preference relation R_e satisfies the Generalized Axiom of Revealed Preference. This approach is significantly more demanding from a computational perspective.

5. A characterization of the efficiency indices

There is a characterization of the set of (e^t) that minimize some norm that will be useful in what follows. In order to describe it, we need some formal definitions.

As above, define the relation R_e^0 by $x^s R_e^0 x^t$ iff $e^t p^t x^t \geq p^t x^s$, and let R_e be the transitive closure of this relation. Then define $GARP_e$ to mean

$$x^s R_e x^t \text{ implies } e^t p^t x^t \leq p^t x^s.$$

If $e^t = 1$ for all t , then this reduces to the standard definition of GARP.

Here is another way to state this definition: if some data (p^t, x^t, e^t) satisfy $GARP_e$, then

$$\text{for all } x^s R_e x^t \text{ we have } e^t p^t x^t \leq p^t x^s.$$

This statement can be written as

$$e^t \leq p^t x^s / p^t x^t \text{ for all } x^s R_e x^t.$$

If we attempt to choose a set of (e^t) that are on the average as close as possible to 1, then this inequality will typically be binding for *some* observations, so we have

$$e^t = \min_{x^s R_e x^t} (p^t x^s / p^t x^t). \quad (2)$$

Note that this is not really an 'operational' way to determine e^t , since e^t is implicitly involved in the relation R_e . Nevertheless, the characterization is still useful, as we shall see shortly.

6. Parametric methods in production analysis

We have seen how to compute measures of goodness-of-fit for nonparametric methods to test models of profit maximization, and utility maximization. However, the same methods can be used in a parametric context.

Consider first the case of profit maximization. Let p^t be the price of output and w^t the vector of factor prices in observation t . Let y^t be the (scalar) output and x^t the vector of factor inputs in observation t . Suppose that we estimate some parametric production function $y = f(x, \beta)$ where β is a vector of parameters. Given an estimate of the parameters, $\hat{\beta}$, we can calculate the maximal profits at each observation t , $\pi(p^t, w^t, \hat{\beta})$. We can then compare the maximal profits from the estimated technology to the actual profits:

$$\pi(p^t, \hat{\beta}) - (p^t y^t - w^t x^t).$$

The magnitude of this number measures the degree to which the observed choice behavior at observation t fails to maximize the *estimated* production function. Hence, it is a measure of how closely the observed production fermion comes to approximating profit-maximizing behavior.

If we are interested in cost-minimizing behavior, we would simply estimate the cost function implied by some parametric production function, $c(w^t, y^t, \hat{\beta})$. The deviation from cost minimization is given by

$$w^t x^t - c(w^t, y^t, \hat{\beta}).$$

This is the difference between the actual cost incurred and the minimal costs, conditional on the assumption that the true technology is of the particular parametric form described by $f(x, \hat{\beta})$.

7. Parametric methods for consumption analysis

Suppose that one is willing to postulate that some observed demand behavior was generated by the maximization of a particular parametric utility function $u(x, \beta)$, where β is a vector of parameters.

Given a parametric utility function $u(x, \beta)$, we can define the associated money metric utility function, $m(p, x, \beta)$, by

$$m(p, x, \beta) = \min_y py,$$

$$\text{s.t. } u(y, \beta) \geq u(x, \beta).$$

In words, the money metric utility function measures the minimum expenditure at prices p the consumer would need to be as well off as he would be consuming the bundle x . For more on the money metric utility function, see Samuelson (1974), King (1982), and Varian (1984).

In terms of the money metric utility function an index of the degree of violation utility-maximizing behavior could be given by

$$i^t = m(p^t, x^t, \beta) / p^t x^t.$$

This index is closely related to the Afriat efficiency index. We can see this by writing i^t as

$$i^t = \min_{u(x, \beta) \geq u(x^t, \beta)} (p^t x / p^t x^t).$$

Note the similarity with eq. (2); the only difference is that e^t uses the partial order over consumption bundles given by the revealed preference relation, while i^t uses the total order over consumption bundles given by the utility function.

This sort of money metric index is a very natural measure of how close the observed consumer choices come to maximizing a particular utility function $u(x, \beta)$. I suggest that it is a useful statistic to report as a goodness-of-fit measure in models of demand estimation.

As with most measures of goodness-of-fit, we can also use the money metric measure as a criterion to estimate the parameters in questions. A natural estimate is to find that value of β that minimizes the degree of violation of maximizing behavior as measured by the values of the indices i^t . For example, one could try to minimize the sum of squares, $\sum_{t=1}^T (i^t)^2$. I believe that this sort of estimator has several desirable properties.

First, it uses a sensible *economic* norm for goodness-of-fit. Conventional estimators of demand parameters use the sum-of-squared errors of the

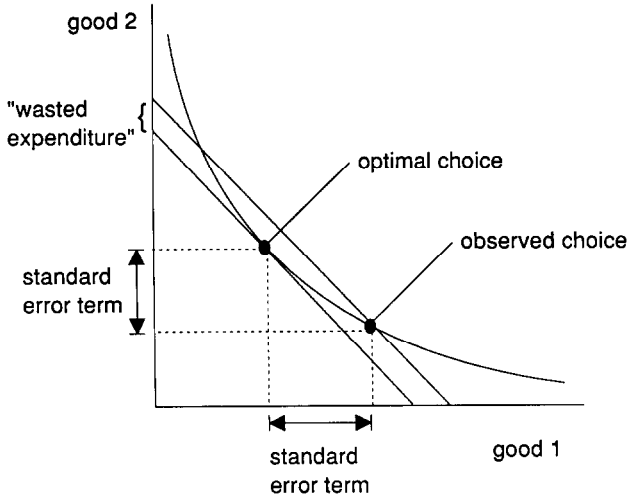


Fig. 1. This is a good fit in terms of money metric utility although it is a bad fit in terms of the usual error terms.

observed and predicted quantities demanded, or some variant on this. But this has little *economic* content, a large difference between predicted and observed demand can easily be consistent with a small difference in utility. This is depicted in fig. 1. Here the observed choice is far from the predicted choice in Euclidean distance, but quite close in terms of money metric utility. The model is a bad fit in terms of Euclidean distance, but a good fit in the sense that the consumer isn't far from maximizing behavior in terms of money metric utility.

Second, the minimized value of the objective function gives a meaningful economic measure of how close the observed choices are to maximizing choice for the particular parametric form involved. If the average value of e^i is 0.95, then it is meaningful to say that the observed choice behavior was 0.95% as efficient as maximizing behavior.

Third, the mechanics of the estimation problem may be much simpler than they are using the conventional approach. Economic theory imposes the restriction that a money metric utility function must be an increasing, linearly homogeneous, and concave function of prices. These constraints are not terribly difficult to impose on the maximization problem. By contrast theory implies that a system of demand equations must have a symmetric negative semidefinite Slutsky substitution matrix. Imposing this restriction involves imposing nonlinear cross-equation restrictions on a system of equations. In general this is a difficult thing to do.

Fourth, this same method can easily be applied to estimation of production relationships. If one starts with a null hypothesis of cost minimization, say, then it makes sense to measure the goodness-of-fit of estimation procedure by comparing the actual costs to the minimum costs implied by the estimated parameters. If it is thought that errors in optimization are a significant component of the error term, then it can make sense to estimate the parameters by choosing parameter estimates that minimize the difference between the observed costs and the minimum costs.

8. An example

In order to examine the money metric goodness-of-fit measure described in the last section, I tried an experiment using U.S. aggregate consumption data. The data were taken from the Citibank economic database and consisted of aggregate consumption of durables, nondurables, and services from 1947 to 1987. The data are presented in table 3.

Like most aggregate consumption data, these figures satisfy the Generalized Axiom of Revealed Preference. This is due to the fact that during the post-war period, most developed economies have experienced reasonably steady real growth: each year has generally been revealed preferred to the previous year and the data trivially satisfy the revealed preference restrictions. Hence the aggregate demand data are consistent with the maximization of utility of a representative consumer.

However, it may be of interest to ask how well common parametric forms of utility functions do in describing these data. It is typically the case that one can reject the restrictions imposed by maximization using parametric forms such as the translog utility function; see, for example, Christensen, Jorgenson, and Lau (1975). But how large are these violations in terms of the *economic* norm described in the last section? In order to answer this question, I estimated the parameters of a Cobb–Douglas utility function and measured the goodness-of-fit using the money metric measure.

The Cobb–Douglas utility function is a convenient parametric form since it has a minimal number of parameters and it automatically satisfies the maximization restrictions. This means that our estimated goodness-of-fit measure will generally be an upper bound on the goodness-of-fit using a more flexible function form. For example, the Cobb–Douglas utility function is a special case of the translog utility function. Thus the ‘best fitting’ Cobb–Douglas function will have at least as good a fit, in terms of our money-metric measure, as the best fitting translog function.

I estimated the parameters of a Cobb–Douglas utility system using three different techniques. The first technique was simply to take the average expenditure share of each good. The second technique was to estimate the

regression $x_i = a_i e / p_i$, where e is the total expenditure on the three goods. I used Zellner's seemingly unrelated regression technique and imposed the normalization that $a_1 + a_2 + a_3 = 1$. (Estimating the three equations separately gave almost the same estimates.) The third technique was to determine the values of the parameters that maximized the goodness-of-fit, as measured by difference between the money metric utility and the actual expenditure. The first two methods are straightforward, but a description of the third method may be in order.

Let us derive the money metric utility function as associated with the Cobb–Douglas utility function $u(x_1, x_2, x_3) = x_1^{a_1} x_2^{a_2} x_3^{a_3}$. For algebraic convenience we impose the normalization that the exponents sum to 1. The money metric utility function is defined to be the amount of money that it takes at some prices (p_1, p_2, p_3) to choose an optimal bundle that has the same utility as the bundle (x_1, x_2, x_3) .

If we let m be the necessary amount of money, we have the equation

$$x_1^{a_1} x_2^{a_2} x_3^{a_3} = (a_1 m / p_1)^{a_1} (a_2 m / p_2)^{a_2} (a_3 m / p_3)^{a_3}.$$

Solving for m we have

$$m(p, x) = a_1^{-a_1} a_2^{-a_2} a_3^{-a_3} (p_1 x_1)^{a_1} (p_2 x_2)^{a_2} (p_3 x_3)^{a_3}. \quad (3)$$

[For a different derivation, see Varian (1984, p. 129).] Taking logs, we can write this equation as

$$\begin{aligned} \ln m(p, x) &= -a_1 \ln a_1 - a_2 \ln a_2 - a_3 \ln a_3 \\ &\quad + a_1 \ln p_1 x_1 + a_2 \ln p_2 x_2 + a_3 \ln p_3 x_3. \end{aligned} \quad (4)$$

We suppose that the log of the actual expenditure in period t , $\ln e^t$, is equal to the log of the expenditure minimizing amount, $\ln m(p^t, x^t)$, plus an error term representing the optimization error. Using eq. (4), we have

$$\begin{aligned} \ln e^t &= -a_1 \ln a_1 - a_2 \ln a_2 - a_3 \ln a_3 \\ &\quad + a_1 \ln p_1^t x_1^t + a_2 \ln p_2^t x_2^t + a_3 \ln p_3^t x_3^t + \varepsilon^t. \end{aligned}$$

I estimated this equation using the nonlinear least squares routine in MicroTSP, imposing the restriction that $a_1 + a_2 + a_3 = 1$. The results from the three estimation methods are in table 1.

The first thing to observe is that the three methods give somewhat different answers. This is simply a consequence of the fact that the estimates which 'fit the data best' depend on what measure of goodness-of-fit you use. The regression estimates that minimizes the sum of squared deviations from the observed demands will not in general be the same as the estimates that

Table 1
Estimated parameter values.

Method	a_1	a_2	a_3
Expenditure shares	0.152	0.461	0.387
Regression	0.129	0.358	0.413
Nonlinear least squares	0.150	0.472	0.378

minimize the squared difference between money metric utility and actual expenditure.

It is surprising that the expenditure share method and the money metric method give very similar estimates, especially since the expenditure share estimate involves a system of equations while the money metric estimation involves only a single equation. Of course, ultimately it is a single sum-of-squares that is minimized in the regression technique, so perhaps this is not so surprising after all.

The computed values of the money metric utility function for each of the different parameters are given in table 2, along with the percentage difference between money metric utility and the actual expenditure for each of the three different estimation methods.

Note that these percent differences are very small, at least for the expenditure share estimates and the NLS estimates. Using the expenditure share methods the largest difference is 7.4%, and the majority of the differences are less than 1%. The average difference is 2%. This suggests that the observed aggregate demand behavior is not very different from optimizing behavior, at least when measured in units of 'wasted expenditure'.⁶

Similar results hold for the nonlinear least squares estimates. Here the average value of the error is only 1.9%. The regression estimates do much poorer, resulting in an average error of about 5%.

It is worth noting that the residuals in all of the estimates are positive in each observation; this is as it should be if the optimizing model is to make any sense since the minimum expenditure to achieve a given level of utility must always be less than an arbitrary expenditure.

A closer examination of the index shows the limitations of the Cobb–Douglas functional form. The Cobb–Douglas form requires that expenditure shares remain constant, while the data clearly show that the share of services in expenditure has significantly increased. Obviously, a more flexible functional form would be appropriate for these data.

⁶Cochrane (1989) independently adopted a similar approach to examining tests of intertemporal consumption models. He finds that the deviation of actual consumption from the optimal intertemporal allocation of consumption is on the order of 30 cents per month – a remarkably small number. Cochrane also discusses the distinction between statistical significance and economic significance in much the same terms as I do.

Table 2
Comparison of estimation techniques.

Year	Actual expenditure	(Shares) m_1	(Regress) m_2	(NLS) m_3	(Shares) $1 - m_1/e$	(Regress) $1 - m_2/e$	(NLS) $1 - m_3/e$
1947	3,855	3,568	3,117	3,600	0.074	0.191	0.066
1948	4,462	4,132	3,609	4,169	0.074	0.191	0.066
1949	4,470	4,235	3,755	4,265	0.053	0.160	0.046
1950	4,876	4,637	4,129	4,666	0.049	0.153	0.043
1951	5,534	5,249	4,658	5,286	0.052	0.158	0.045
1952	5,872	5,634	5,057	5,670	0.041	0.139	0.034
1953	6,257	6,088	5,543	6,120	0.027	0.114	0.022
1954	6,576	6,444	5,924	6,474	0.020	0.099	0.015
1955	7,285	7,149	6,585	7,177	0.019	0.096	0.015
1956	7,816	7,713	7,163	7,740	0.013	0.084	0.010
1957	8,537	8,441	7,867	8,469	0.011	0.079	0.008
1958	9,048	8,979	8,436	9,006	0.008	0.068	0.005
1959	9,939	9,881	9,318	9,904	0.006	0.063	0.004
1960	10,560	10,525	10,002	10,545	0.003	0.053	0.001
1961	11,019	10,996	10,517	11,014	0.002	0.046	0.000
1962	11,933	11,915	11,423	11,930	0.002	0.043	0.000
1963	12,785	12,767	12,290	12,777	0.001	0.039	0.001
1964	13,914	13,894	13,414	13,901	0.001	0.036	0.001
1965	15,350	15,319	14,793	15,325	0.002	0.036	0.002
1966	17,163	17,135	16,550	17,143	0.002	0.036	0.001
1967	18,581	18,554	18,038	18,555	0.001	0.029	0.001
1968	21,431	21,380	20,804	21,376	0.002	0.029	0.003
1969	24,340	24,282	23,729	24,272	0.002	0.025	0.003
1970	27,471	27,389	26,961	27,372	0.003	0.019	0.004
1971	31,294	31,124	30,829	31,082	0.005	0.015	0.007
1972	35,846	35,591	35,382	35,531	0.007	0.013	0.009
1973	42,530	42,374	41,812	42,341	0.004	0.017	0.004
1974	50,710	50,507	49,743	50,503	0.004	0.019	0.004
1975	61,383	61,131	60,316	61,113	0.004	0.017	0.004
1976	72,478	72,090	71,506	72,012	0.005	0.013	0.006
1977	85,701	84,969	84,889	84,822	0.009	0.009	0.010
1978	102,261	101,237	101,465	101,043	0.010	0.008	0.012
1979	124,104	122,988	123,047	122,804	0.009	0.009	0.010
1980	148,952	146,795	148,177	146,525	0.014	0.005	0.016
1981	180,406	177,153	179,803	176,757	0.018	0.003	0.020
1982	205,088	199,545	204,941	198,908	0.027	0.001	0.030
1983	232,873	225,046	232,866	224,161	0.034	0.000	0.037
1984	263,642	253,883	263,598	252,772	0.037	0.000	0.041
1985	294,808	281,220	294,461	279,813	0.046	0.001	0.051
1986	322,488	304,152	321,354	302,408	0.057	0.004	0.062
1987	353,848	331,047	351,875	329,083	0.064	0.006	0.070
				Mean	0.020	0.052	0.019

To the extent that a more flexible functional form would fit the data better than the Cobb–Douglas form, our goodness-of-fit measure should be regarded as an upper bound on ‘wasted expenditure’. If the average ‘wasted expenditure’ in the Cobb–Douglas case is less than 2%, it would be even less if we used a more flexible functional form.

Table 3
U.S. Aggregate consumption (Citibank economic database).

Year	Durables p_1	Nondurables p_2	Services p_3	Durables x_1	Nondurables x_2	Services x_3
1947	27.40	26.88	16.85	20.43	90.88	50.60
1948	28.77	29.18	17.77	22.85	96.60	55.48
1949	29.95	27.85	18.45	25.05	94.85	58.42
1950	31.23	27.73	18.90	30.73	98.22	63.15
1951	31.85	29.73	19.40	29.85	109.15	69.03
1952	31.57	29.85	20.30	29.23	114.72	75.13
1953	31.77	29.38	21.40	32.67	117.83	82.13
1954	32.10	30.02	22.18	32.10	119.67	88.05
1955	33.38	30.60	23.02	38.88	124.70	94.30
1956	34.73	30.95	24.02	38.20	130.78	101.63
1957	36.50	31.90	25.02	39.65	137.10	108.55
1958	36.92	32.85	26.10	37.17	141.75	115.67
1959	38.50	33.17	26.93	42.80	148.47	125.00
1960	38.77	33.63	27.80	43.42	153.20	134.00
1961	38.88	34.08	28.40	41.90	157.40	141.80
1962	39.75	34.58	29.13	47.02	163.82	151.05
1963	40.15	34.90	29.85	51.80	169.35	160.63
1964	40.35	35.25	30.60	56.85	179.68	172.78
1965	40.92	36.10	31.43	63.48	191.85	185.40
1966	41.55	37.35	32.60	68.53	208.45	200.30
1967	42.33	38.23	33.80	70.63	216.90	216.00
1968	44.55	39.98	35.65	81.00	235.00	236.43
1969	46.10	41.98	37.70	86.22	252.18	259.43
1970	47.80	44.05	40.38	85.67	270.32	284.02
1971	50.25	45.92	43.08	97.58	283.27	310.65
1972	51.05	47.90	45.58	111.22	305.10	341.27
1973	50.83	53.75	48.10	124.72	339.55	372.98
1974	54.08	59.58	51.77	123.75	380.90	411.90
1975	61.15	65.13	56.38	135.35	416.20	461.23
1976	65.88	67.60	60.65	161.45	451.95	515.92
1977	69.15	71.03	65.45	184.50	490.45	582.25
1978	72.78	76.00	70.30	205.57	541.80	656.10
1979	78.25	83.55	75.88	218.95	613.25	734.55
1980	85.55	88.85	83.72	219.28	681.35	831.95
1981	93.65	96.78	92.30	239.88	740.58	934.70
1982	100.00	100.00	100.03	252.65	771.00	1026.97
1983	101.60	102.50	106.13	289.10	816.70	1128.75
1984	102.97	106.17	111.60	335.55	867.30	1227.63
1985	102.35	108.50	117.25	368.70	913.13	1347.52
1986	101.38	110.13	122.25	402.43	939.35	1458.05
1987	100.40	114.05	127.42	413.73	982.88	1571.22

References

- Afriat, S., 1967, The construction of a utility function from expenditure data, *International Economic Review*, 8, 67-77.
- Afriat, S., 1972, Efficiency estimates of production functions, *International Economic Review*, 8, 568-598.
- Akerlof, G. and J. Yellen, 1985, Can small derivations from rationality make significant differences to economic equilibria?, *American Economic Review* 75, 708-720.

- Banker, R. and A. Maindiratta, 1988, Nonparametric analysis of technical and allocative efficiencies in production, *Econometrica* 56, 1315–1332.
- Bronars, S., 1987, The power of nonparametric tests, *Econometrica* 55, 693–698.
- Browning, M., 1984, A non-parametric test of the life-cycle rational expectations hypothesis, *International Economic Review* 30, 979–992.
- Christensen, L., D. Jorgensen, and L. Lau, 1975, Transcendental logarithmic utility functions, *American Economic Review* 65, 367–383.
- Cochrane, J., 1989, The sensitivity of tests of intertemporal allocation of consumption to near-rational alternatives, *American Economic Review* 79, 319–337.
- Deaton, A., 1987, Life-cycle models of consumption: Is the evidence consistent with the theory?, in: T. Bewley, ed., *Advances in econometrics – Fifth world congress*, Vol. II (Cambridge University Press, Cambridge).
- Diewert, E. and C. Parkan, 1985, Tests for consistency of consumer data and nonparametric index numbers, *Journal of Econometrics* 30, 127–147.
- Green, R. and S. Srivastava, 1985, Risk aversion and arbitrage, *Journal of Finance* 40, 257–268.
- Green, R. and S. Srivastava, 1986, Expected utility maximization and demand behavior, *Journal of Economic Theory* 38, 313–323.
- Hanoch, G. and M. Rothschild, 1972, Testing the assumptions of production theory: A nonparametric approach, *Journal of Political Economy* 8, 256–272.
- Houtman, M. and J. Maks, 1987, The existence of homothetic utility functions generating Dutch consumer data (University of Groningen, Groningen).
- King, M., 1982, Welfare analysis of tax reforms using household data, *Journal of Public Economics* 21, 183–214.
- Landsburg, S., 1981, Taste change in the United Kingdom, 1900–1955, *Journal of Political Economy* 89, 92–104.
- Manser, M. and R. McDonald, 1988, An analysis of substitution bias in measuring inflation, 1959–1985, *Econometrica* 56, 909–930.
- McCloskey, D., 1985, The loss function has been mislaid: The rhetoric of significance tests, *American Economic Review* 75, 201–205.
- McCloskey, D., 1989, Formalism in economics, rhetorically speaking, *Ricerche Economiche* 43, 57–75.
- Samuelson, P., 1938, A note on the pure theory of consumer behavior, *Economica* 5, 61–71.
- Samuelson, P., 1947, *Foundations of economic analysis* (Harvard University Press, Cambridge, MA).
- Samuelson, P., 1974, Complementarity: An essay on the 40th anniversary of the Hicks–Allen revolution in demand theory, *Journal of Economic Literature* 12, 1255–1289.
- Varian, H., 1982a, The nonparametric approach to demand analysis, *Econometrica* 50, 945–972.
- Varian, H., 1982b, Nonparametric test of models of consumer behavior, *Review of Economic Studies* 50, 99–110.
- Varian, H., 1984, *Microeconomic analysis* (W.W. Norton, New York, NY).
- Varian, H., 1984, The nonparametric approach to production analysis, *Econometrica* 52, 579–597.
- Varian, H., 1985, Nonparametric analysis of optimizing behavior with measurement error, *Journal of Econometrics* 30, 445–458.
- Varian, H., 1988, Revealed preference with a subset of goods, *Journal of Economic Theory* 46, 179–185.