Because energy conservation programs are generally voluntary, participating households are different from nonparticipants in important, energy-related ways. This self-selection bias complicates efforts to estimate energy savings due to these programs. This article discusses several methods for dealing with self-selection. The choices include nonrandom sampling of program nonparticipants, binary choice models that explicitly treat household decisions to participate and to retrofit, or use of both methods. Because some of the methods discussed are new and have not yet been applied to analysis of energy conservation programs, we developed a "synthetic" data set. We conducted numerical experiments with this data to examine the performance of these different methods. These experiments show that the improved sample design and analytical techniques generally yield more accurate estimates of program energy savings. Our experience also suggests that a small, well-defined synthetic data set is helpful in developing, debugging, and evaluating software associated with new analytical approaches.

USE OF SYNTHETIC DATA IN DEALING WITH SELF-SELECTION

Improving Conservation Program Energy Savings Estimates

ERIC HIRST
JOHN TRIMBLE
RICHARD GOELTZ
N. SCOTT CARDELL
Energy Division,
Oak Ridge National Laboratory

Government agencies and utilities offer a variety of energy conservation services to their citizens and to their customers. During the past few years, the scope, size, and cost of these programs have

AUTHORS' NOTE: *Research sponsored jointly by the Bonneville Power administration and the Office of Conservation and Renewable Energy, U.S. Department of Energy, under contract W-7405-eng-26 with Union Carbide Corporation.*

EVALUATION REVIEW, Vol. 7 No. 6, December 1983 807-830

increased substantially (Willrich and Kubitz, 1981; California Energy Commission, 1980). As the size and cost of these programs increase, it becomes increasingly important that efforts be made to carefully measure the effects of these programs (Office of Technology Assessment, 1980; Office of Environmental Engineering, 1981): both the extent to which these programs reduce energy consumption, and the cost-effectiveness of the programs to participating households, nonparticipating households, utilities, and society in general, see Soderstrom et al., 1981; Hirst et al., 1982.

Developing reliable estimates of program energy savings and cost-effectiveness is a complicated and subtle task. One of the major complications is "self-selection." Because these conservation programs are generally voluntary, the households that choose to participate in the programs are almost certain to be different from nonparticipating households in important, energy-related ways. Failure to account for this self-selection bias in evaluation design and analysis will lead to energy-saving estimates that include the effects of both the program and of self-selection.

The self-selection process has two aspects. First, energy consumption and program participation are likely to be interdependent. High energy users are more likely to participate in programs (because they have more to gain from participation) than are low energy users. Participation, in turn, is likely to lead to lower energy consumption. This simultaneity problem, if not properly addressed, will yield an underestimate of program energy savings.

Second, households that are interested in conservation are more likely to participate than are households that are either indifferent or opposed to conservation. Failure to correct for this bias will yield an overestimate of program energy savings.

This paper discusses methods to deal with self-selection in evaluation of residential energy conservation programs. Because some of the methods have not yet been applied to analysis of conservation programs, we were unsure about their feasibility in time and dollar costs to implement, and what they contribute to greater accuracy and reliability of energy-saving estimates. Therefore, we developed a "synthetic" data set with which we tested a variety of methods for estimating program energy savings.

Section 2 provides additional information concerning self-selection in residential conservation programs, and Sections 3 and 4 present alternative ways of dealing with this problem. Section 5 presents the synthetic data and discusses how and why we created this data set. Section 6 discusses results obtained with the data set. The first section discusses the relative strengths and weaknesses of different ways to treat self-selection.[1]

SECTION 2—THE SELF-SELECTION PROBLEM

The typical conservation program is offered to all eligible[2] households; that is, the utility or government agency does not determine by random assignment or requirement who participates. For example, a utility may offer its residential customers a free home energy audit. The purpose of the audit is to inspect the dwelling unit and the energy-using equipment within the unit and, based on the auditor's observations, to recommend energy conservation practices and measures suitable for that particular residence.[3] Some utilities offer financial incentives (low- or zero-interest loans, or cash rebates) to encourage installation of the suggested conservation measures.

Because participation in these programs is voluntary, a self-selection bias may be present. Therefore, one must determine whether observed changes in energy use are due to the program itself rather than the composition of the group that accepts the program. Evidence shows that conservation program participants typically do not represent a cross-section of the general public: audit program participants generally have higher education and income levels than nonparticipants, are more aware of energy problems, own larger homes, and consume more energy than nonparticipants (Berry et al., 1981; Berry, 1982; Newman and Day, 1975). As a consequence, comparison of energy savings between program participants and nonparticipants will not yield an accurate estimate of program energy savings. The energy savings achieved by participants relative to nonparticipants is due to both the program and the self-selection process.

The ideal method to estimate program effects is to use random assignment (Cook and Campbell, 1979). For example, in the early stages of program implementation, the utility or government agency might receive many more requests for program services, such as energy audits, than its limited resources can accommodate. Randomly assigning people who request services to either receive the service or receive a placebo service, like energy conservation publications, creates two comparable groups. Comparisons can then be made between the two groups without concern for self-selection, because both groups selected themselves into the program.

Unfortunately, very few utilities and government agencies take advantage of the start-up phase of a program or use pilot programs to design and conduct careful evaluations.[4] As a result, most evaluation is done post-hoc; the evaluator is faced with a program that operates throughout the region that is included in the agency's political jurisdiction or the utility's service area. So, a search must be made for alternative approaches to estimate program effects. The problem can be defined by considering four groups, shown in the schematic below: those who participate in the program (A), those who do not participate in the program (C), and the behaviors of both groups without the program (B and D). Evaluation seeks to measure the differences between energy-related behaviors with and without the program (A-B and C-D). However, because the program exists, conditions B and D cannot be observed.

	PARTICIPANT	NONPARTICIPANT
PROGRAM	A	C
NO PROGRAM	B	D

Random assignment allows one to create the B group.[5] Without such ideal circumstances, the evaluator can draw a sample from the nonparticipants (C) to approximate what the participants would have done without the program (B) or the evaluator can conduct ex post analyses to infer from either A or C or both, the behavior of B; both notions are discussed below.

Some proposals to deal with self-selection involve using the audit group as its own control, purposely withholding program services from some requestors, or using sequential sampling to ascertain changes in participants as programs progress. Using the audit group as its own control which is sometimes called an interrupted time series design; see Cook and Campbell, 1979, is difficult because it requires considerable data for both the pre- and post-program time periods. In addition, the validity of this design rests on the assumption that no other factors were operating during this time that would substantially affect energy use; yet other conservation programs and changes in weather, fuel prices, and incomes are likely to affect energy use.[6] Thus, this approach does not appear feasible for evaluation of conservation programs.

Purposely withholding program services from some requestors suffers from two problems. First, withholding services may be politically unwise or prohibited by regulation. Second, households that know they will be provided program services at a later date may not behave as if the program did not exist. On the contrary, they are likely to purposely defer conservation actions until program services are provided. Thus, use of such a comparison group is likely to yield an overestimate of program effects.

The third approach, comparing households that participate in the program at different times, is feasible (Weiss and Newcomb, 1981) and was suggested as part of our evaluation plan for the Bonneville Power Administration. This is shown in T1 and T2 of Table 1; also, see Hirst et al., 1982). Such a comparison also allows the evaluator to see how the characteristics of program participants change over time.

This design includes four separate groups of households. This allows multiple approaches to estimation of program effects, a kind of "triangulation." Although each approach has limitations, if the results from each are roughly comparable, an evaluator has much greater confidence in the overall estimates of program effects.

SECTION 3—SAMPLE DESIGN TO ASSIST IN ANALYSIS

We propose an intentional nonrandom sample of nonparticipants (to keep sampling costs low) matched to participants on the basis of preprogram energy use; nonandomness due to self-selection, which is participation, is then explicitly modeled in analyzing the data. This process is made more statistically efficient by matching a sample of nonparticipant households to a sample of participants on the basis of a stratified sampling scheme that uses residential, appliance, demographic, and attitudinal characteristics (Trimble, 1982; Hausman and Trimble, 1981).

Data from nonparticipants (C1 in Table 1) can be used to develop a matched subsample. One approach is to use direct matching on explanatory variables, which unfortunately suffers from three problems. The

TABLE 1
Overall Evaluation Design for the Bonneville Power Administration

Groups	*Dates* (7/1980)	*1980/1981*		*1981/1982*		*1982/1983*	*(6/1983)*
T1		U	P	U	m,t	U	
T2		U		U	P,m,t	U	
C1		U		U	m	U	
C2		U		U	m,t	U	

T1— sample of households that participated in Spring 1981.
T2— sample of households that participated in Spring 1982.
C1— sample of households randomly selected from utility customer lists.
C2— sample of nonparticipant households selected from C1 to form a subsample "matched" with T1 and/or T2.
U— utility bills (consumption, price) and matched weather data for winter heating months (e.g., October through March); also other data (e.g., zip code, rate class, heating fuel, years at present address) readily available from utility account records.
P— program participation (audit, loan, water heater wrap, etc.); restricted to services *between* heating seasons.
m— mail screener survey conducted among 1981 and 1982 program participants and random sample of utility customers.
t— telephone survey conducted among 1981 and 1982 program participants and a matched sample of nonparticipants.
SOURCE: (Hirst et al., 1982)

first is a practical one, related to the computational complexity of matching on more than a few variables. Second, direct matching gives equal weights to the matching variables, which may not accurately reflect the relative importance of these variables. Third, large sample sizes are required to match with more than a few variables.

Because of these problems, we suggest an alternative scheme (Hausman and Trimble, 1981; Cochran, 1963; Manski and McFadden, 1981) based on the assumption that energy consumption is a key determinant of program participation.[7] In this scheme, a regression equation is estimated to explain variation in energy consumption as a function of the variables collected in a screener survey or that is available from utility records. This equation is then used to define "predicted" energy consumption for nonparticipant households. Nonparticipants are then assigned to strata based on their predicted preprogram energy use. Random samples are drawn from each stratum such that the distribution of predicted energy consumption among this subsample of nonparticipants (C2) is similar to the distribution of actual energy consumption among program participants (T1 and T2).

Matching on predicted energy consumption will improve the statistical efficiency of the estimation process, discussed in the following section, by reducing the variance of the estimators. The success of the match depends on the explanatory power (R^2) of the regression model of energy use; which, in turn, is affected by the extent to which the important determinants of energy use are accurately captured in the screener survey. Also, the matching scheme is based on the assumption that program participation and preprogram energy use are correlated. This is because households with high fuel bills are likely to benefit most from a conservation program. To the extent that other factors not related to energy use also influence participation, the power of the matching procedure is reduced.[8]

SECTION 4—ANALYSIS TO DEAL WITH SELF-SELECTION

Estimation of program energy savings (ES) involves the following:

$$\begin{aligned} \text{ES} &= \text{Expected} \left(\begin{matrix}\text{Participant}\\ \text{energy use}\end{matrix} \,\middle|\, \text{without program}\right) \\ &= \text{Expected} \left(\begin{matrix}\text{Participant}\\ \text{energy use}\end{matrix} \,\middle|\, \text{with program}\right). \end{aligned}$$

The second term can be directly estimated with the data. The first term must be inferred, based on theory and data. Complications occur because energy use, participation, and retrofit are all interdependent.

The simplest way to begin analysis is to compute means (and standard deviations) for the groups of households (Table 1) for the pre- and post-program time periods (Table 2). These values can be compared to see whether the change in mean energy use is greater for the program groups (T1, T2) than for the nonprogram groups (C1, C2).

Ignoring nonparticipants and comparing pre- and post-program energy use for participants only (T1, T2) shows how much energy was saved by participants (Ozenne and Reisner, 1980). However, unless the participant and nonparticipant groups are equivalent, as would be the case with random assignment, and no factors other than the particular program affected energy use between time periods 1 and 2, this method does not yield a reliable estimate of program energy savings. Rather, the result shows the combined effect of the program, self-selection, and other factos that affect energy use (such as a change in weather or increases in fuel prices).

A second method involves a comparison of means between post-program consumption of participants and nonparticipants. This result is a reliable estimate of program energy savings only if the two groups are equivalent before the program begins, in other words, if there are no energy-related pre-program differences between the two groups. This assumption is also unlikely to be correct. In addition, this method assumes that nonprogram factors affect participant and nonparticipant energy use in the same ways.

A third method compares changes in energy use between the two groups (Dukich, 1982). While this approach is better than either of the preceding ones, it still assumes that the two groups are equivalent and that other factors that influence energy use affect participants and nonparticipants in the same way.

A fourth method is to construct regression models for energy consumption (Parti and Parti, 1980; Taylor, 1975; Hannigan and King, 1982; and Grady and Hirst, 1982). This explicitly accounts for various determinants of household energy use (demographic characteristics, structure characteristics, weather, fuel prices; as well as program participation). A single equation can be developed in which all the households for both time periods are pooled:

$$E_{it} = f(X_{it},P_{it}),$$

where i refers to individual households and t refers to time period (pre-or post-program). P is a dummy variable for program participation; P is always zero for nonparticipants and is 1.0 for program participants for the months after retrofit. The magnitude and statistical significance of the coefficient of this dummy variable indicate the energy-saving effect of the program. Unfortunately, this coefficient also includes the effect of self-selection. X is a vector of factors such as heating degree days and household income that influence energy use. Use of the matched sub-sample of nonparticipants (C2) in this analysis reduces the variances of the parameter estimates and the likelihood of colinearity among explanatory variables, relative to an analysis conducted with a simple random sample of nonparticipants, C1.

A variation on this approach allows the coefficients—the sensitivity of energy use to different factors—to vary across the participant and nonparticipant groups and across the two time periods (Burnett, 1982). The preceding equation is expanded to:

$$E = f(X, X{*}D_t, X{*}P, X{*}D_t{*}P)$$

where D_t is a dummy variable for year, equal to zero for the pre-program period and equal to one for the post-program period. The coefficients of the last term in the equation are used to estimate the program's energy saving.

A more sophisticated approach (5, in Table 2) is to develop a system of equations, one to model program participation and a second to model energy use (Williams and Walther, 1982; Dubin, 1982; Dubin and McFadden, 1982; Henson, 1982; Olsen, 1980; Heckman, 1978, 1979; Goett and McFadden, 1982). The equation to predict the probability of program participation uses a logit or probit formulation (because participation is a discrete choice) with predicted pre-program energy use (E^{1*}, or its determinants) and economic, demographic, and structural factors as explanatory variables.

The model of energy use is similar to that described above except that a function of the predicted probability of participation, from the first equation, is used to adjust the regression equation. There are several ways to specify the probability-of-participation model, the adjusting function, and the energy demand model. The present method uses the Mills ratio, defined below (Dubin and McFadden, 1982; Olsen, 1980; Heckman, 1979). With the adjustment provided by the Mills ratio, one can properly specify a model with a dummy variable for participation.

The binomial choice model for program participation is:

$$P = g(X, W, E^{1*}).$$

Predicted values of probability of program participation (P^*) are then used to define a Mills ratio for participants and nonparticipants:

$$MR_T = +\ln(P^*) + [(1 - P^*)/P^*]\ln(1 - P^*) \qquad \text{(participants)}$$

and

$$MR_C = -\ln(1 - P^*) - [P^*/(1 - P^*)]\ln(P^*). \quad \text{(nonparticipants)}$$

TABLE 2
Summary of Methods to Estimate Program Energy Savings

Description	*Method*[a]	*Remarks*
1. Comparison of means, participants only	$E_T^1 - E_T^1$	Simplest approach; assumes nonparticipants make *no* energy-use changes or that participants would have made no energy-use changes without the program[b]
2. Comparison of post-program means	$E_C^1 - E_T^1$	Simple approach; assumes no pre-program differences between groups[b]
3. Comparison of means across groups and time periods	$(E_T^1 - E_T^2) - (E_C^1 - E_C^2)$	Simple approach; assumes two groups the same[b]
4. Simple regression equation	$E = f(X,P)$	Accounts for other factors (weather, income, fuel prices) that influence energy use in addition to program participation[b]
5. Explicit prediction of participation and energy use	$P = g(X,W,E^{1*})$ $E = f(X,P^*)$	Logit or probit formulation used to account for self-selection by modeling participation; correlation between participation and energy use accounted for
6. Explicit prediction of participation, retrofit and energy use	$R = h(X,Z,E^{1*})$ $P = g(X,W,E^{1*},R^*)$ $E = f(X,R^*,P^*)$	Logit or probit used to model participation and retrofit decisions; both retrofit and participation treated explicitly in energy equation

a. The symbols used are:

E is monthly or annual energy use per household; E* is predicted energy use,
1 refers to the preprogram period; 2 refers to the postprogram period,
T is the treatment (program) group; C is the comparison group,
X is a vector of factors that influence energy use,
W is a vector of factors unique to program participation,
Z is a vector of factors unique to retrofit,
P is a dummy variable for program participation; P* is the Mills ratio, based on predicted probability of participation,
R is a dummy variable for retrofit (installation of conservation measures); R* is based on the predicted probability of retrofit.

b. These methods do not correct for self-selection and are useful primarily for exploratory data analysis.

SOURCES: (Hirst et al., 1982 and 1983)

The Mills ratio is then used in the energy demand equation to purge the other coefficients of the self-selection bias:

$$E = f(X,P,MR)$$

where MR_T is used for participants and MR_C for nonparticipants.

This approach has two advantages. First, it corrects for the simultaneity problem with respect to participation and usage. The earlier suggestion to use a dummy variable in the energy-use equation assumes that participation is independent of usage. This approach recognizes the interaction between these two factors. Second, this approach allows an explicit analysis of the factors that influence program participation, an important issue in program management and operation.

A closely related technique is to use the values of predicted participation (from the binary choice model of participation) as instrumental variables in place of the participation dummy variable in the energy demand equation. While estimation of the energy demand model uses the predicted probability of participation for each household, simulation with the energy demand model, to estimate program energy savings, uses the actual (0,1) values for participation. This technique can also be used with the last approach, discussed below.

The final approach (6, in Table 2) that we considered involves explicit and interdependent estimation of the decisions to participate in the program and retrofit (in a nested logit formulation), and energy use. Here again, there are several ways to specify the models; we again choose a logit formulation to model each of the four alternatives: retrofit/participate, retrofit/not participate, not retrofit/participate, and not rerofit/not participate. A nested logit formulation is specified to take account of the correlations among the retrofit and participation decisions. In this case, binary choice models are first developed for the retrofit decision, with separate models estimated for program participants and nonparticipants:

$$R_j = h_j(X,Z,E^{1*}) \qquad j = 1,2.$$

These two equations are used to calculate inclusive values (IV) for participants and nonparticipants

$$IV_j = -\ln(1 - R_j^*) \qquad j = 1,2,$$

where R^* is the predicted probability of retrofit computed from the appropriate retrofit choice model. These inclusive values take account of the correlation between the retrofit/participation decisions for participants and nonparticipants.

Next, a choice model for participation is estimated:

$$P = g(X, W, E^{1*}, R, IV_1, IV_2),$$

where the IV values are defined above.

Finally, a regression model for energy use is estimated as follows:

$$E = f(X, P^*R, P^*NR, NP^*R, NP^*NR, MR).$$

P and R are participation and retrofit dummy variables; NP and NR are nonparticipation $(1 - P)$ and no retrofit $(1 - R)$ dummy variables. MR is the Mills ratio, compute as before.

This approach ensures that energy demand is explicitly responsive to both participation and retrofit and, through the Mills ratio, corrects for the self-selection bias discussed previously.

One can envision further extensions to this model. For example, multinomial choice models can be developed to estimate the probability of choosing individual retrofit measures, or "packages" of measures. A simplification of this approach would use retrofit cost and expected energy savings for each household's retrofit package in the binomial retrofit choice equation discussed above. Models can be developed to predict changes in energy consumption $(E_1 - E_2)$ or to predict post-program energy use conditional on pre-program energy use, such as $E_2 = f(X,P,E_1)$. All of the models discussed in this section can be developed with either or both groups of participants (T1 and T2, from Table 1) and with either the random or matched sample of nonparticipants (C1 or C2).

SECTION 5—SYNTHETIC DATA

The preceding section discussed methods to adjust for self-selection in estimating the energy savings due to conservation programs. These methods have not yet been employed in analysis of energy conservation programs. Therefore, we developed a synthetic data set to explore possible problems associated with implementation of these ideas and the improvement in analytical results that they yield.

Use of synthetic data provides several attractions relative to use of actual program data. First, by creating the data the analyst knows the "correct" answer; that is, actual program energy saving is known. Second, one can modify the data to determine how robust various analytical approaches are to changes in the data which include the particular random sample drawn, the size of the sample, the influence of the program on retrofit, the effect of retrofit on energy savings, and the strength of the self-selection bias. Finally, the synthetic data can be created to focus on the particular issue of interest. For example, in the present case, our data assume that energy use is independent of fuel prices and of weather.

The major drawback to analysis conducted with such data is uncertainty over the extent to which the synthetic data reflect accurately the behavior of households. This uncertainty makes it difficult to generalize from results obtained with synthetic data to what might happen in evaluation of a real program.

In the data discussed below, we assume that post-program energy use is a function only of pre-program energy use, retrofit, and program participation (as well as the usual unobservable effects reflected in the error terms). In actuality, post-program energy use is affected by changes in weather, fuel prices, household size, structure size and additions or demolitions, appliance holdings, and other factors. If any of these factors is correlated with program participation, retrofit, or pre-program energy use, the actual analysis becomes much more complicated than indicated in this study.

The synthetic data we created include the following variables: pre-program energy use (E1), participation in the hypothetical conservation program (P), retrofit (R), and post-program energy use (E2). These variables are assumed to be functions of socioeconomic and dwelling unit (structure) characteristics: income (Y), number of years in present home (YEAR), floor area of home (FT2), and age of home (HAGE) are used as proxies for these characteristics in our data set. The equations used in creating this data are:

$$E1 = a_1 + a_2 FT2 + a_3 Y + a_4 e_w + a_5 e_z + a_6 e_{E1}$$

$$Z = b_1 + b_2 E1 + b_3/(1 + YEAR) + b_4 e_Z$$
$$P = 1 \text{ if } Z > 0, \text{ else } P = 0$$

$$W = c_1 + c_2 E1 + c_3 HAGE + c_4 P + c_5 e_W$$
$$R = 1 \text{ if } W > 0, \text{ else } R = 0$$

$$E2 = E1 - k(d_1 P + d_2 P{*}E1 + d_3 R + d_4 R{*}E1 + d_5 e_W + d_6 e_Z + a_6 e_{E1})$$

where e_i are random error terms intended to reflect errors in the data and equation form, and the effects of variables not included in the equations, and a_i, b_i, c_i, d_i, and k are coefficients determined by the analyst.

Values for Y, FT2, YEAR, and HAGE are drawn randomly from lognormal distributions with means and standard deviations determined by the analyst and assigned to households in the synthetic data set. Energy use is assumed to be a function of household income and dwelling unit floor area, as proxies for socioeconomic and structure characteristics, respectively. In addition, energy use is influenced by independent random errors, with components related to retrofit, participation, and energy use. These random components all have zero means and standard deviations proportional to their coefficients; e_{E1} is normally distributed, while e_Z and e_W are logistically distributed.

Z is an index of program participation, assumed proportional to pre-program energy use and inversely proportional to the number of years the household has lived in the dwelling unit. The coefficient of E1 (b_1) determines the extent of self-selection in program participation, as discussed above. If Z is greater than zero, the household is a program participant (P = 1); otherwise the household does not participate in the program (P = 0).

Similarly, W is an index of household retrofit (installation of conservation measures to reduce energy consumption). W is assumed proportional to pre-program energy use and to the age of the dwelling unit. In addition, W is positively related to participation as reflected by the c_4^P term in the equation.

Finally, post-program energy use is assumed equal to pre-program energy use, minus energy savings, plus error terms related to W, Z, and E1. Energy use is reduced through two mechanisms. First, households that participate in the program may save energy regardless of whether they retrofit. Such savings could occur through adoption of energy conservation practices such as lowering thermostat settings on heating and water heating systems recommended in the home energy audit. Second, households can reduce energy use by installation of retrofit measures. Thus program participation can affect post-program energy use and energy savings in two ways: directly through adoption of conservation practices and indirectly (through the W equation) through stimulation of installation of retrofit measures. Finally, post-program energy use is affected by random error terms similar to those that influence preprogram energy use (e_Z, e_W, e_{E1}).

Although the data set described above is much simpler than what might actually occur, it is still surprisingly complicated. In particular, there are a total of 23 coefficients to adjust.

SECTION 6—RESULTS OBTAINED WITH THE SYNTHETIC DATA

We used the synthetic data to examine the performance of the matching procedure (Section 3) and the modeling approaches (Section 4) that include estimation of program participation and retrofit. Because of the very low cost associated with creation and use of the synthetic data, we were able to examine the following kinds of cases in a sensitivity analysis:

(1) Variations in sample size (500, 1000, 2500 households).
(2) Variations in the sample drawn from the total population (different initial random seeds).
(3) Effects of the participation and retrofit error terms on preprogram energy use (a_4, a_5).
(4) Effects of the participation and retrofit error terms on post-program energy use (d_5 and d_6).
(5) Effects of the energy use error term (a_6) on pre- and post-program energy use.
(6) Effects of pre-program energy use on participation (b_2).
(7) Effects of participation on energy savings

directly (d_1 and d_2); and

indirectly (c_4, d_3, and d_4).

(8) Effects of practices and retrofit on energy savings (k).

We compared results obtained with different procedures in terms of their ability to predict program energy savings (ES). The "true" value of ES was computed as follows. First, the data were created as outlined in the preceding section. Then, the equations for W (retrofit) and E2 were solved again for program participants, setting c_4, d_1, and d_2 all equal to zero, which would assume that the program did not exist. The true value of ES is then ES from the first step minus ES from the second step (the difference between participant energy savings with the program and energy savings for the same participants without the program).

The baseline case used in this discussion involved creation of 1000 households as shown by Table 3. Households are assigned to the

TABLE 3
Mean Values for Key Variables in the Base Case[a]

	Participants	*Nonparticipants*	
		Total	*Matched sample*
Number	119	881	119
Income (thousand $)	16.8	14.7	16.7
Floor area (ft^2)	1400	1230	1490
Years in present home (years)	5.4	7.1	7.7
Age of dwelling unit (years)	24.6	24.3	23.2
Preprogram energy use (MBtu)	257	189	244
Fraction that retrofit	0.83	0.27	0.32
Postprogram energy use (MBtu)	195	179	229
Difference between pre- and post-program energy use (MBtu)	62	10	15
Participant behavior without the program			
Postprogram energy use (MBtu)	235		
Difference between pre- and post-program energy use (MBtu)	22		

a. Coefficients in this base case (Section 5) are:

a_i: -150, 0.16, 10, 3, 3, 40
b_i: -7.3, 0.013, 4, 1
c_i: -4.63, 0.006, 0.1, 1.5, 1
d_i: 20, 0.1, 30, 0.15, 0, 0
k : 0.8

conservation program on the basis of their preprogram energy use and years of occupancy in their present home as noted by the coefficients in footnote 'a' of Table 3. As a consequence, average energy use for participants is 36% higher than for nonparticipants. Participants reduce their energy consumption by an average of 62MBtu; nonparticipants cut consumption by only 10 MBtu.

The matching procedure described in Section 3 yields a sample of nonparticipants that is similar to the participants. This is noted in Table 3. In particular, preprogram energy use for the matched nonparticipants is much closer to that for participants than for nonparticipants in general. Also, the matched nonparticipants are much more like the participants than nonparticipants in general in terms of floor area and income, the determinants of preprogram energy use. On the other hand, the matched nonparticipants are much more like the nonparticipants in general in terms of years of residence (a determinant of program participation but not of energy use).

It is interesting to compare the performance of the matching procedure and the analytical techniques shown in Table 2, and described in Section 4 with the "true" program energy saving [40 MBtu (= 62 – 22 from Table 3)]; as shown in Table 4. The comparisons using this baseline show several interesting features. There is considerable variation in estimated energy savings across the different procedures. The first method, based only on changes in consumption for participants, consistently overestimates program effects. The second method, based only on postprogram energy consumption, consistently underestimates program effects (and sometimes, as in the baseline, yields negative program energy savings).

The third method, comparison of mean energy savings of participants and nonparticipants, often gives reasonably good results; this is surprising given the simplicity of this method.

The models that explicitly treat program participation and use the Mills ratio to adjust for self-selection (#5 and #6 in Table 4) generally yield more accurate predictions than do the simpler regression models (#4). In particular, the models that explicitly treat participation, retrofit, and energy use (#6) almost always yield more accurate answers than do the other approaches.

Analyses using the matched sample of nonparticipants rather than a random sample of nonparticipants generally yield more accurate estimates of program energy saving. The one consistent exception to this generality concerns the simple regression model (#4), for which results are often better without matching.

All but two of the program energy-savings estimates shown in Table 4 are statistically significant at the 1% level or better. The exceptions are method 2, simple comparison of post-program consumption levels for the full sample and the random subsample. The Mills ratio terms in models 5 and 6 are all highly significant, confirming the self-selection built into the data.

Many other cases were run using this data set as perturbations around the baseline. Although there were differences among runs in which methods worked best, there was surprising consistency among the results (Hurst et al., 1983). We tested data sets with a different a_6 error term in the E1 equation to increase or decrease the standard deviation of E1 and E2. We tested cases in which the influence of the error terms from participation and retrofit (a_4 and a_5) on E1 were changed. Changing these error terms has a much larger effect on the simpler models (#1 through #4, Table 4) than on the two models that use the Mills ratio

TABLE 4
Energy Savings Estimated with Different Methods and Data Sets[a]
(Ratio of Estimated-to-Actual Energy Saving)

	Data set		
	All data (n = 1000)	*Subsample of nonparticipants*	
		matched (n = 238)	*random (n = 222)*
1. [b]$E_i^1 - E_i^2$	1.56	1.56	1.56
2. $E_c^2 - E_i^2$	−0.40[c]	0.84	−0.29[c]
3. $(E_i^1 - E_i^2) - (E_c^1 - E_c^2)$	1.30	1.19	1.28
4. E = f(Y, FT2, P, DYR)	0.69	0.53	0.70
5. P = g(Y, FT2, YEAR)			
E = f(Y, FT2, P, DYR, MR)	1.26	1.17	1.27
6. R = h(Y, FT2, HAGE)			
P = g(Y, FT2, YEAR, R, IV)	1.11	0.99	1.06
E = f(Y, FT2, P*R, P*NR, NP*R, NP*NR, DYR, MR)			

a. "True" program energy saving is 40 MBtu. Baseline run.
b. These numbers refer to the models listed in Table 2.
c. Not statistically significant.

correction. These results suggest that the sophisticated models are robust with respect to variations in the error terms, which are random variations in E1 and their relationship to participation and retrofit.

We tested cases in which the determinants of program participation (b_1) are changed. One case involved random assignment to the program ($b_2 = b_3 = 0$) and the other involved an increase in the effect of preprogram energy use on participation (b_2). With random participation in the program, the models with Mills ratio corrections (#5 and #6, Table 4) yielded less accurate results than did the simpler models. On the other hand, the matching procedure yielded better results with every model. When the self-protection bias associated with program participation was increased, the models with the Mills ratio adjustment performed substantially better than did the models without this correction for self-selection. The matching procedure yielded more accurate results in all cases, except for Model 4.

We examined cases in which the participation and retrofit error terms in the E2 equation were varied. We tested cases in which the number of observations (n), the initial seed, the effect of program participation on retrofit (c_4), the effect of participation on energy saving (d_1 and d_2), the effect of retrofit energy saving (d_3 and d_4), and the net effect of participation and retrofit on energy saving (k) were varied. To avoid repetition of similar results we report only the final case, in which overall energy savings are reduced.

We tested a case with k = 0.95 rather than 0.8 as in the baseline to see how well the models perform when the overall program energy savings are reduced from 40 MBtu in the baseline to 21 MBtu. This is a stringent test presented in Table 5 of these analytical approaches because the actual program energy saving is only 10% of preprogram energy consumption.

As with the other cases discussed above, the first two methods yield results that are very high (model 1) or very low (model 2). The third model, the difference between mean energy savings of participants and nonparticipants, yields estimates that are 20 to 30% higher than the actual value. The coefficient of the participation dummy variable in model 4 was statistically insignificant with each of the three data sets. The models with the Mills ratio corrections perform very well.

SECTION 7—SUMMARY

This article discusses several methods for dealing with self-selection in energy conservation programs for which participation is voluntary. These methods involve nonrandom sampling of program nonparticipants, binary choice models that explicitly treat decisions to participate and to retrofit, or both.

Because these methods are new and have not yet been applied to evaluation of conservation programs, we developed a "synthetic" data set. We used this data set to develop and debug software associated with these methods. We then conducted numerical experiments to examine the performance of these methods in accurately predicting program energy savings.

These numerical experiments lead to a number of conclusions. First, the matching procedure, in which a nonrandom sample of nonparticipants is matched to participants on the basis of predicted preprogram energy use, generally leads to more accurate predictions of program energy savings. A frequent exception to this generality are the simple regression models, #4 in Table 2. In the other cases, however, both the

TABLE 5
Energy Savings Estimated with Different Methods and Data Sets[a]
(Ratio of Estimated-to-Actual Energy Savings)

	Data Set		
	All data ((n = 1000)	*Subsample of nonparticipants*	
		matched (n = 238)	*random (N =222)*
1. $E_T^1 - E_T^2$	1.56	1.56	1.56
2. $E_C^2 - E_T^2$	-1.97	0.21[b]	-1.75
3. $(E_T^1 - E_T^2) - (E_C^1 - E_C^2)$	1.30	1.19	1.28
4. E = f(Y, FT2, P, DYR)	0.15[b]	-0.09[b]	0.19[b]
5. P = g(Y, FT2, YEAR) E' F(Y, FT2, P, DYR, MR)	1.21	1.14	1.25
6. R = h(Y, FT2, HAGE) P = g(Y, FT2, YEAR, R, IV) E = f(Y, FT2, P*R, P*NR, NP*R, NP*NR, DYR, MR)	1.00	0.79	0.90

a. "True" program energy saving is 21 MBtu. Baseline run with k = 0.95.
b. Not statistically significant.

simple comparisons of means and the more sophisticated Mills ratio models, matching improves accuracy of prediction.

Second, the models that explicitly treat the decisions to participate in the program and also to retrofit generally yield more accurate estimates of program energy savings than do the simpler regression models. In addition to this obvious advantage, the qualitative choice models of the decisions to participate and to retrofit are useful in their own right. These models provide a quantitative explanation of the factors (and their importance) that affect the decisions to participate and to retrofit. This information should be helpful to conservation program managers in terms of forecasting future program effects and in improving program marketing strategies.

The choice models are valuable because they more closely reflect the actual decisonmaking processes of households than do the simpler energy demand models. As a consequence, these more sophisticated approaches are useful in explaining changes in energy use, while the simpler models are useful primarily in describing, but not explaining these changes.

NOTES

1. It is important to note that many other problems complicate use of actual fuel bills to estimate conservation program energy savings (Hirst, 1981). These include changes across both time and location in weather, fuel prices, household structure, and demographic characteristics. Errors in the data and missing data elements further complicate such analyses.

2. Eligibility requirements might include restrictions to certain housing types, primary heating fuels, and income groups.

3. The federal Residential Conservation Service, created by the 1978 National Energy Conservation Policy Act, requires major gas and electric utilities to offer such programs to their residential customers. (U.S. Congress, 1978; U.S. Department of Energy, 1982).

4. See Johnston (1982) for one of the rare instances in which random assignment was successfully used in an energy conservation program.

5. In an actual program, nonparticipants (the C group) may be indirectly affected by the program through, for example, discussions with neighbors who participated in the program; we assume that this second-order effect is small.

6. If sufficient time series data are available, these factors can be (at least partially) controlled for in a multivariate regression model of energy use.

7. For example, our ongoing evaluations of residential weatherization programs in Minnesota and the Pacific Northwest show that preprogram energy use is 20 to 30% higher for participants than for nonparticipants.

8. An alternative although similar approach is to develop a nonrandom sample of nonparticipants using predicted program participation as the matching criterion, instead of preprogram energy use. In this case, a qualitative choice model, such as the logit or probit model), is developed to predict participation; see methods 5 and 6 in Table 2. A sample of nonparticipants is drawn that has a distribution of predicted participation similar to that for participants (Williams and Walther, 1982). This method is not explored here.

REFERENCES

BERRY, L., J. SODERSTROM, and E. HIRST (1981) Review of Evaluations of Utility Home Energy Audit Programs, Oak Ridge National Laboratory, ORNL/CON-58, March.

BERRY, L. (1982) "Characteristics of Participants in Home Energy Audit Programs," Chapter 4 in Evaluation Plan for the Bonneville Power Administration Residential Energy Conservation Programs, Volume II, Oak Ridge National Laboratory, ORNL/CON-94, April.

BRADLEY, R., and J. S. SHAFFER (1981) Evaluation of Seattle City Light's Residential Insulation Program, Seattle City Light, February.

BURNETT, T. (1982) "Measuring Weatherization Effectiveness: Portland General Electric Company's Experience," Proceedings of the EPRI Workshop on Measuring the Effects of Utility Conservation Programs, Electric Power Research Institute, EPRI EA-2496, July.

California Energy Commission (1980) Energy Service Corporations: Opportunities for California Utilities, December.

COCHRAN, W. G. (1963) Sampling Techniques, New York: John Wiley.

COOK, T. D., and D. T. CAMPBELL (1979) Quasi-Experimentation, Design & Analysis Issues for Field Settings, Skokie, IL: Rand McNally.

DUBIN, J. A. (1982) Economic Theory and Estimation of the Demand for Consumer Durable Goods and Their Utilization: Appliance Choice and the Demand for Electricity, Massachusetts Institute of Technology, Energy Laboratory, MIT-EL-82-035WP, May.

DUBIN J. A., and McFADDEN (1982) "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," Proceedings: End-Use Models and Conservation Analysis, Electric Power Research Institute, EPRI EA-2509, July.

DUKICH, T. D. (1982) Testimony before the Idaho Public Utilities Commission (Case No. P-300-24), Washington Water Power Company, June.

GOETT, A., and D. McFADDEN (1982) Residential End-Use Energy Planning System (REEPS), Cambridge Systematics, Inc., prepared for the Electric Power Research Institute, EPRI EA-2512, July.

GRADY, S., and E. HIRST (1982) Evaluation of Utility Home Energy Audit Programs: A Wisconsin Example, Oak Ridge National Laboratory, ORNL/CON-88, March.

HANNIGAN, S. R., and P. KING (1982) "Residential Conservation Programs at Pacific Power and Light Company: Models, Forecasts, and Assessments," Proceedings of the EPRI Workshop on Measuring the Effects of Utility Conservation Programs, Electric Power Research Institute, EPRI EA-2496, July.

HAUSMAN, J. A., and J. L. TRIMBLE (1981) "Sample Design Considerations and Sample Selection for the Vermont Time-of-Day Electricity Use Survey," Review of Public Data Use 9(4): 83-96.

HECKMAN, J. J. (1978) "Dummy Endogenous Variables in a Simultaneous Equation System" Econometrica 46 (July).

HECKMAN, J. J. (1979) "Sample Selection Bias as a Specification Error," Econometrica 47 (January).

HENSON, S. E. (1982) An Econometric Analysis of the Residential Demand for Electricity in the Pacific Northwest, Department of Economics, University of Oregon, July.

HIRST, E. (1981) "Measuring the Effects of Utility Conservation Programs," Public Utilities Fortnightly 108 December 17.

HIRST, E., et al. (1982) Evaluation Plan for the Bonneville Power Administration Residential Energy Conservation Program Vol. 1, Oak Ridge National Laboratory, ORNL/CON-94, July.

HIRST, E., et al. (1983) Methods to Deal with Self-Selection in Estimating Conservation Program Energy Savings: Use of Synthetic Data to Test Alternative Approaches, Oak Ridge National Laboratory, ORNL/CON-120, February.

JOHNSTON, R. (1982) Residential Conservation Service Audits: A Behavioral Evaluation of Four Types, Southern California Edison Company, August.

MANSKI, C., and D. McFADDEN (1981) "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in Structural Analysis of Discrete Data with Econometric Applications, C. Manski and D. McFadden (eds.), MIT Press.

NEWMAN, D. and D. DAY (1975) The American Energy Consumer, Cambridge, MA: Ballinger.

Office of Environmental Engineering (1981) Review of the Department of Energy's Conservation and Solar Energy Programs, U.S. Environmental Protection Agency, EPA-600/7-81-01, January.

Office of Technology Assessment (1980) Conservation and Solar Energy Programs of the Department of Energy: A Critique, U.S. Congress, U.S. Government Printing Office.

OLSEN, R. J. (1980) "A Least Squares Correction for Selectivity Bias," Econometrica 48 (November).

OZENNE, D. C., and R. A. F. REISNER (1980) The TVA Home Insulation Program: An Evaluation of Early Program Impact, ICF, Inc., prepared for the Tennessee Valley Authority, April.

PARTI, M., and C. PARTI (1980) "The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector," Bell Journal of Economics 11 (Spring): 309-321.

SODERSTROM, E. J., et al. (1981) Evaluation of Conservation Programs: A Primer, Oak Ridge National Laboratory, ORNL/CON-76, July.

TAYLOR, L. D. (1975) "The Demand for Electricity: A Survey," Bell Journal of Economics 6 (Spring): 74-110.

TRIMBLE, J. L. (1982) "Sample Selection Issues and Methods for Evaluation Studies in Natural Environments," Chapter 5 in Evaluation Plan for the Bonneville Power Administration Residential Energy Conservation Programs, Volume II, Oak Ridge National Laboratory, ORNL/CON-94, April.

U. S. Congress (1978) National Energy Conservation Policy Act, Public Law 95-619, November 9.

U.S. Department of Energy (1982) "Residential Conservation Service Program, Final Rule," Federal Register 47, U.S. Government Printing Office, June 25.

WEISS, C. S., and T. M. NEWCOMB (1981) Evaluation of the Home Energy Check Program, Seattle City Light, October.

WILLIAMS, M. V., and R. J. WALTHER (1982) "Issues in the Evaluation of Conservation Programs: Nonrandom Data and Participation Bias," Proceedings of the EPRI Workshop on *Measuring the Effects of Utility Conservation Programs,* Electric Power Research Institute, EPRI EA-2496, July.

WILLRICH, M., and K. KUBITZ (1981) "Organization and Regulatory Alternatives for the Energy Utility Industry," Public Utilities Fortnightly 107, June 18.

Eric Hirst has been with the Oak Ridge National Laboratory since 1970. His research deals with engineering, economic and policy issues connected with energy use and conservation in the residential and commercial sectors. He was on assignment with the Minnesota Energy Agency during 1979, and earlier spent 15 months as Director of the Office of Transportation Research in the Federal Energy Administration. Dr. Hirst holds a Ph.D. in Mechanical Engineering from Stanford University and has to his credit more than 100 publications on energy and environmental issues.

John Trimble has been with the Oak Ridge National Laboratory since 1978 doing research in applied econometrics and energy demand. He was formerly a visiting faculty member at *the University of Colorado at Boulder. Dr. Trimble holds a Ph.D. in economics from Texas A & M University and has published several articles applying econometric techniques to address energy demand issues.*

Richard Goeltz has been a computer specialist for Oak Ridge National Laboratory since 1980. He has a B.A. degree in computer science and mathematics from the University of Tennessee. Mr. Goeltz provides programming, data-base design, and data analysis for a multidisciplinary research team at Oak Ridge.

Nicholas Scott Cardell is an econometrician who studied under Zvi Griliches at Harvard University. He invented the Nested Logit and Hedonic Demand models. He has been employed on a full-time basis by Charles River Associates and Oak Ridge National Laboratory, and on a consulting basis by National Economic Research Associates, the National Academy of Sciences and Abt Associates.