WORKSHOP

An energy utility was required by its regulator to administer a program to promote energy conservation by businesses in its service territory. Under this program, utility representatives inspect commercial and industrial firms' premises and suggest ways in which each firm can reduce its energy consumption. These inspections are called audits, and the actions that are suggested include operational changes (such as turning out the lights when everyone leaves the building) and capital expenditures (such as replacing an inefficient air conditioning system with a more efficient one). The audited firm does not have to take any of the suggested actions. However, the recommended actions are supposedly cost-effective for the firm (in that the energy savings from the actions would reduce the firm's energy bill sufficiently to warrant the cost of the actions), and consequently it is expected that audited firms will take at least some of the recommended actions.

An important question in evaluating this program is whether the audits actually result in energy savings, and, if so, what is the magnitude of the savings. By comparing the savings from the program with the cost of the program, an informed decision can be made as to whether the program should be continued.

To investigate this issue, data were collected on a random sample of 500 firms in the utility's territory. For each firm, the following data were obtained:

USAGE	=	The firm's average monthly electricity consumption in 1992 in kilowatthours (kWh).
AC	=	A dummy variable indicating whether the firm had been audited prior to 1992 (that is, AC = 1 if the firm had been audited, AC = 0 if the firm had not been audited).
SQFT	=	The size of the firm's premises, measured in thousands of square feet.
NEMPLOY	=	The number of employees who work on the firm's premises.
CDD	=	A measure of the weather in the firm's location in 1992. This variable is called "cooling degree days" and is traditionally used to capture the extent to which the firm needs to air condition its premises. In hot climates, CDD is higher than in cooler climates. You do not really need to know the exact definition, but if you are interested it is the following: For each day the extent to which the average temperature of the day exceeds 65 degrees is recorded (if average temperature is below 65, then zero is recorded). The CDD for the year is this amount summed over all days in the year.

Each month, the utility selected firms to be audited in that month, selecting firm randomly from its list of customer accounts (omitting those already audited). Thus, whether a firm was audited or not prior to 1992 is independent of the actions of the firm.

Also, firms that were audited during 1992 were eliminated from the sample. Consequently, the sample consists of firms that were audited before 1992 or will eventually be audited (if the program continues) after 1992.

Your goal is to estimate the effect of an audit on firms' energy consumption. This information will be used to evaluate the audit program and to decide whether the money spent on the audits is warranted.

1. Why is it necessary to eliminate firms that were audited in 1992?

2. The data are contained in an ASCII file in the **tue** directory. You need to read these data into TSP and create a TSP data file before you run any regressions in TSP. The file is named DATASET.ASC.

   The file DATASET.ASC contains 500 lines of data, one per firm. Each line contains five fields: one field for AC, CDD, NEMPLOY, SQFT, and USAGE, in that order. The file is in "free format," meaning that each field is separated by a blank space.

(i) Look at the first lines of the file by typing:

```
more DATASET.ASC        <cr>
```

This command allows you to view the contents of the file one screenful at a time. To continue scrolling through the file, press the space bar. To quit, type q or ^c (the CTRL + c keys together).

(ii) TSP allows you to read data into TSP in a variety of ways through the LOAD command. For our purposes, create the file containing the following commands that will read your data and create a TSP data file called FIRMDAT--which is the same data as in DATASET.ASC but in TSP format instead of ASCII. (The TSP datafile is easier for TSP to use.)

```
options crt;
smpl 1 500;
out firmdat;
load (file = "DATASET.ASC") ac, cdd, nemploy,
       sqft, usage;
out;
msd ac, cdd, nemploy, sqft, usage;
```

The OUT command creates the TSP data file **firmdat** and puts into it all variables loaded or created until it comes to a second OUT command.

After running this program, you can, in future TSP programs, simply say

```
in firmdat;
```

to get your data.

3. Run a regression of USAGE against AC, CDD, NEMPLOY, SQFT, and an intercept.

   (i) What is the estimated savings due to the audit?

   (ii) Divide the estimated savings by the average USAGE to get savings as a proportion of average consumption. Does the result seem reasonable.?

   (iii) We expect electricity consumption to be greater for large buildings (as measured by SQFT) and for buildings in hotter areas (as measured by CDD). Are these expectations confirmed in your regression? Another factor that we expect to affect electricity consumption in a building is more people. People within a building generate heat such that more air conditioning is required (and hence more electricity would be consumed) in buildings with lots of people in them. In light of this, why do you think NEMPLOY is entering with a negative coefficient?

   (iv) Do you think the R-squared is reasonable?

4. Try a log-linear form of the regression. That is, take the log of USAGE, SQFT, NEMPLOY, and CDD. (The log of AC cannot be taken because AC = 0 for many firms and the log of zero is undefined. In any case, the log of a dummy is a dummy, such that entering AC or log AC would produce the same results.) Regress the log of USAGE against an intercept, AC, and the logs of SQFT, NEMPLOY, and CDD.

   The TSP command to create a log of a variable, say log of USAGE, is:

   ```
   genr lnu = log(usage);
   ```

   where **lnu** is the new variable name for the log of USAGE.

   (i) Does taking logs increase or decrease the R^2?

   (ii) How do you interpret the coefficient of AC?

   Recall: log USAGE = a + bAC + ... is the same as USAGE = exp (a + bAC + ...).

   So:

   $$\frac{\partial \text{USAGE}}{\partial \text{AC}} = \exp(a + bAC + ...) \cdot b = b\,\text{USAGE}$$

   Does the estimated coefficient of AC seem reasonable? How does it compare to your result in question 3(ii)?

5. It might be more reasonable to consider electricity consumption per square foot rather than total electricity consumption. Estimate a model with the dependent variable being USAGE/SQFT and the explanatory variables being a constant, AC, NEMPLOY, and CDD.

   Do not take logs: this model will be compared with the model from step 3.

   (i) What does this model assume about the relation between electricity consumption and the size of a building?

   (ii) Does this model seem to fit the data better or worse than the model in step 3?

   (iii) Interpret the estimated coefficient of AC. What is the average estimated savings from an audit (i.e., averaged over all firms?). How does this estimate compare to that obtained in step 3(i)?

6. The utility supposedly selected firms to be audited in a random fashion.

   (i) If the utility actually chose firms to be audited first that it knew or expected would be more likely to implement the recommended actions, would our regression model provide an unbiased estimate of the average savings from an audit? Explain, including a description of the direction of bias if it exists.

   (ii) If firms were allowed to call the utility and request an audit, and the utility audited these firms shortly after they called, would our estimation results be biased? If so, in what direction? Explain.