WORKSHOP

1. Consider the first model you estimated in the previous workshop. The output is reproduced below:

```
Dependent variable: USAGE
Current sample:  1 to 500
Number of observations:  500

Mean of dependent variable = 229970.
Std. dev. of dependent var. = 91538.8
Sum of squared residuals = .362272E+13
Variance of residuals = .731863E+10
Std. error of regression = 85549.0
R-squared = .133589
Adjusted R-squared = .126588
Durbin-Watson statistic = 2.06459
F-statistic (zero slopes) = 19.0806
Schwarz Bayes. Info. Crit. = 22.7658
Log of likelihood function = -6385.38

           Estimated     Standard
Variable   Coefficient     Error       t-statistic
C          127568.       15133.6         8.42949
AC         -20606.7      8116.66        -2.53882
CDD        23.9860       4.80534         4.99152
NEMPLOY    -6.61867      89.4999        -.073952
SQFT       931.616       147.957         6.29652
```

a. Using the t-statistic, test the hypothesis that the coefficient of CDD is zero. Do you reject the hypothesis or not?

b. Using the standard error, find the 95% confidence interval for the coefficient of CDD. Does this interval overlap zero?

c. Repeat steps (a) and (b) for the coefficient NEMPLOY.

d. Another researcher estimated that audits reduce firms' energy consumption by 25,000 kWh per month on average. Test the hypothesis that the true impact of an audit is 25,000. The test is performed as follows.

- The hypothesized value of β, denoted β_H, is 25,000.

- Your point estimate of β, denoted $\hat{\beta}$, is 20,606.7 with a standard error, s, of 8116.7.

- The t-statistic for this hypothesis is

$$\frac{\hat{\beta} - \beta_H}{s} = \frac{20{,}606.7 - 25{,}000}{8116.7}$$
$$= -0.54$$

- The critical value of the t-statistic at 95% confidence is ±1.96 (essentially, ±2). Because the t-statistic for your hypothesis is below 1.96 in magnitude, you do **not** reject the hypothesis.

  Alternatively, you can calculate the P-value. In TSP,

  CDF(T,DF=495) -0.54

gives .58944, which is bigger than 0.05.

Why is this t-statistic (-0.54) different from the t-statistic given in the output (-2.54)?

The test indicates that you cannot reject the hypothesis that the true impact of the audit is 25,000 kWh. Does this mean that the true impact ***is*** 25,000 kWh?

e. The people who run the audit program have reported that the audits reduce customers' consumption by 40,000 kWh monthly on average. Using the same steps as for part d, test the hypothesis that the true savings is 40,000. Can you reject or not reject this hypothesis?

f. For what range of values can you not reject the hypothesis that the true impact of the audit is that value?

g. What is your best estimate of the impact of the audit?

2. If the hypothesis that a coefficient is zero cannot be rejected, then the variable is said to be "insignificant" or to "enter insignificantly." And if the hypothesis that the coefficient is zero ***can*** be rejected, the variable is said to be "significant" or to "enter significantly."

   Some researchers routinely omit variables that are insignificant and leave only significant variables. This practice is ***sometimes*** advisable, but sometimes it is not a reasonable practice.

   a. Rerun the model from part 1 with insignificant variables removed. Do you think this is a better or worse model than the original one? (There is no right answer here--it depends on your own views.)

   b. Now suppose that you had estimated your model in log-linear form. This was the model you estimated in step 4 of the previous workshop. The output is given below:

```
Dependent variable: LU
Current sample:  1 to 500
Number of observations:  500

 Mean of dependent variable = 12.2558

         Adjusted R-squared = .066134
Std. dev. of dependent var. = .458604
    Durbin-Watson statistic = 2.07536
   Sum of squared residuals = 97.2224
  F-statistic (zero slopes) = 9.83443
       Variance of residuals = .196409
 Schwarz Bayes. Info. Crit. = -1.57546
   Std. error of regression = .443180
 Log of likelihood function = -300.067
                  R-squared = .073620

           Estimated    Standard
Variable  Coefficient    Error      t-statistic
C         10.9689       .252391        43.4599
AC        -.017402      .041390        -.420451
LC        .093483       .026117        3.57941
LN        -.519830E-02  .033586        -.154777
LS        .151217       .031652        4.77753
```

Which variables are insignificant in this model? Do you think the insignificant variables should be dropped? What do you think is the reason for AC entering insignificantly in this model when it entered significantly in the previous model?

c. Let's go back to the linear model. There are good reasons for believing that the impact of an audit is larger for larger buildings. (Why?) To capture this effect, create a new variable and enter it into the model; that is, run these TSP commands:

```
in firmdat;
genr acs = ac*sqft;
olsq usage c,ac,acs,cdd,sqft,nemploy;
```

(i) Does the coefficient of **acs** confirm your expectations?

(ii) What is the estimated impact of an audit on average? (Recall that the average of **sqft** is 70.4.) Compare this with the estimate obtained in the model with **ac** only (not **acs**).

(iii) Both **ac** and **acs** are insignificant in this model. Why is **ac** insignificant now when it was significant in the original model?

3. Sometimes a significant variable should be dropped. Run a model with the dependent variable being usage per square foot and all the explanatory variables **ac**, **sqft**, **cdd**, **nemploy**, and a constant:

```
genr ups = usage/sqft;
olsq ups c,ac,sqft,cdd,nemploy;
```

Note which variables are significant and which are not. What is causing this pattern of results?

4. Test whether there is evidence that the variable **sqft** does not account for all differences between small and large buildings. Estimate an unrestricted regression function in which buildings with **sqft** greater than 70 have different coefficients than smaller buildings and compute an F test for whether the coefficients are equal.

   (a) To estimate the unrestricted regression, create variables for **ac**, **sqft**, **cdd**, **nemploy** and the constant equal to a dummy for observations with **sqft** greater than 70 times each explanatory variable.

   ```
   genr c1 = sqft>70;
   genr ac1 = c1*ac;
   genr cdd1 = c1*cdd;
   genr sqft1 = c1*sqft;
   genr nemploy1 = c1*nemploy;
   ```

   (b) Compute the unrestricted regression by including these variables as additional explanatory variables in the regression of usage on a constant, **ac**, **cdd**, **sqft**, and **nemploy**. Why is this specification equivalent to different coefficients for the large buildings? How do you interpret the coefficients for **c1** through **nemploy1**?

   (c) Use the restricted and unrestricted sum of squared residuals statistics, and the appropriate degrees of freedom to compute the F-statistic for equal coefficients. What do you conclude?