Problem Set #4

ECONOMICS 141
FALL 2008

Due November 6

1. Again using the data on fuel economy and automobile speed in question #1 of problem set #2, use the TSP (or alternative) statistical package to verify the "stepwise regression" results for the quadratic term in the quadratic regression model of question #2, problem set #3. Specifically, you should first estimate (again) the regression coefficients of fuel economy F_i on a constant, speed S_i, and its squared value $(S_i)^2$; then regress both F_i and $(S_i)^2$ on a constant and S_i, and regress the residuals for F_i on the residuals for $(S_i)^2$. Verify the equality of the two estimates of the coefficient on $(S_i)^2$ and report the value of R^2 for the two regressions. Also, show that the reported standard error for the "stepwise" regression coefficient on $(S_i)^2$ is smaller than the corresponding standard error for the "long" regression by a factor of $\sqrt{[(N-K)/(N-1)]}$, where N is the sample size ($N = 12$ here) and K is the number of regressors in the "long" regression ($K = 3$).

2. Suppose you are interested in monthly expenditures on public transportation by individual households, and believe that the logarithm of public transit expenses (denoted $TRANS$) is linearly related to the log of the local price of transit ($PTRANS$), the log of the price of gasoline ($PGAS$), the log of family income ($INCOME$), and the log of the number of licensed drivers in the household ($DRIVERS$). Given an sample of $N = 35$ households, the following least-squares results are available:

Variable	Estimate	Standard Error
Constant	1.2	0.4
PTRANS	-0.5	0.3
PGAS	0.2	0.3
INCOME	0.7	0.2
DRIVERS	-0.1	0.2

Estimated error variance : $s_\varepsilon^2 = 0.09.$
Multiple correlation coefficient : $R^2 = 0.40.$

(a) Suppose the data are generated from a classical regression model with normal errors, and that you can treat the observed regressors as fixed. Determine which of the coefficient estimates are (individually) significantly different from zero at a 5% level. Also, test the null hypothesis $H_0 : \beta_4 = 1$ at a 5% level, where β_4 is the (true) coefficient on log income.

(b) Again under the same assumptions as part (a), test the null hypothesis that all coefficients except the intercept term are zero, again at a 5% level.

(c) Now suppose the sample of 35 observations was grouped into two nonoverlapping groups (of 15 low and 20 high income households), and that for the first (low income) group the sum of squared errors is $SSE = 0.5$. while for the second group the residual sum of squares is $SSE = 1.0$. Use an F-test to test the hypothesis of no difference in regression coefficients across groups, with the usual 5% significance level.

(d) If the squared values of the residuals $\hat{e}_i^2$ from the estimated equation were regressed against the levels of the explanatory variables (including a constant) , and if the R^2 from this second-step regression is 0.3, would you reject the null hypothesis of homoskedastic disturbances at an approximate 5% significance level?

(e) Using the same data set, suppose a "restricted regression" was estimated, using the log of transportation expenses per driver, $TRANS - DRIVERS$, as the dependent variable, with regressors being a constant term, the log of the relative prices of transit to gasoline (i.e., $PTRANS - PGAS$), and the log of income per driver ($INCOME - DRIVERS$). If you used the the sum of squares from this regression, along with the original regression results, to construct an F-test, what would be the degrees of freedom for this test? In terms of the five true coefficients β_1 through β_5 for the unrestricted model (where β_1 is the intercept term, β_2 is the coefficient on the $PTRANS$ variable, etc.), what null hypothesis on the β coefficients could be tested using this F statistic?

[REMARK: Questions #2 appeared on the midterm exam for an econometrics course comparable to this one, and may be a useful example for the next midterm.]

3. **(Due November 18)** A famous problem in economics (studied by Ernst Engel) is whether food expenditure is proportionally related to income. Defining

$$\begin{aligned} Y_i &= \text{logarithm of food expenditure for household } i\text{, and} \\ X_i &= \text{logarithm of income for household } i, \end{aligned}$$

the null hypothesis of proportionality is equivalent to the null hypothesis that, if Y_i and X_i are assumed to follow the standard (bivariate) linear model, then the true value of the regression coefficient for X_i equals one.

1. (a) Some of Engel's original data for English working-class families (circa 1900) are available in the data file "Engel.dat" on the course website. Use the 235 observations on EXP (food expenditure) and INC (income) to test the null hypothesis of proportionality at the 5% level, assuming the error terms are i.i.d. and normally distributed. (Remember to take logs first.)

   (b) Carry out the same test as in part (a), but use the Eicker-White robust standard errors (using the TSP option "(ROBUSTSE)" for the OLSQ command) instead of the usual normal-theory estimator. Do your conclusions change?

   (c) Regress the squared values of the residuals against a constant term, X_i and X_i^2 to test the null hypothesis of homoskedastic error terms at an (approximate) significance level of 5%.

   (d) Use the inverses of the predicted values from the "squared residual regression" in part (c) above as weights in a weighted least squares regression of Y_i on X_i, (with the "(WEIGHT=W)" option in TSP) and perform the same test of proportionality as in (a), assuming that the quadratic specification for heteroskedasticity in part (c) above is correct. Do the conclusions from part (a) above change?