Econometrics 2 – Regression, ctd.

Thursday, March 22th, 2015

1. (Greene) **Regression fit.** The NewGas.dta file contains an updated version of the gasoline market data for 1953-2004. The variables in this data set are:

   YEAR = 1953 … 2004
   GasExp = index number for gasoline expenditure
   GasQ = index number for gasoline physical consumption
   GasP = index number for gasoline price
   GasCPIU = index number for gasoline component for urban clerical worker's CPI
   PCIncome = per capita real income
   PNC = price index for new cars
   PUC = price index for used cars
   PPT = price index for public transportation
   PD = Price index for durables component of total consumption
   PN = Price index for nondurables component of total consumption PS =
   Price index for services component of total consumption
   Pop = population in thousands

   Using these 52 observations on the gasoline data compute the following:

   a) Compute the linear regression of *GasQ/Pop* on a constant, *GasP*, *PCIncome*, *PN*, *PD*, *PS*, and *YEAR*-1952, Report the coefficients, the R^2, and the total, regression, and residual sums of squares.
   b) Compute the linear regression of log*GasQ/Pop* on a constant, log*GasP*, log*PCIncome*, log*PN*, log*PD*, log*PS*, and *YEAR*-1952 (not the log of the trend). Report the same computations as in a). Which model fits better? Explain.

2. (Greene) Estimate by least squares a modified version of the regression model of part 1. b) which looks as follows:

$$\log GasQ/Pop = \beta_1 + \beta_2(\log GasP \text{ up to 1973, 0 else}) + \beta_3(\log GasP \text{ after 1973, 0 else}) +$$
$$\beta_4 \log PCIncome + \beta_5(YEAR - 1952) +$$

   Now, use least squares to fit the coefficients of the model

$$\log GasQ/Pop = \beta_1 + \beta_2 \log GasP + \beta_3(\log GasP \text{ after 1973, 0 else})$$
$$+ \beta_4 \log PCIncome + \beta_5(YEAR - 1952) +$$

   Report the least squares coefficients for both cases. Could you have computed the second least squares regression from the first one?

3. (Greene) **Regression without a constant term.** What is the effect on R^2 (or, sum of squared residuals) of computing a linear regression without a constant term?

4. (Greene) In the loglinear regression in 1. b), a (questionable) microeconomic theory might predict that the three elasticities on nondurables price, log*PN*, durables price, log*PD*, and services price, log*PS*, should sum to one. A certain professor also thinks that the time trend should not be present. Compute the **constrained regression** which imposes these four restrictions. Once again, compute the R^2 and the components of the analysis of variance. Compare these results to those in 1. b).

5. (Greene) The following exercises use the gasoline data which we have used before. All variables in all regressions discussed below are assumed to be logarithms. As such, in the results to follow, all estimated coefficients are estimates of elasticities.
    a) The correlation of log*GasQPC* and log*GasP* is positive. In the regression of log*GasQPC* on a constant term, log*GasP* and log*PCIncome*, the slope on log*GasP* is negative. Obtain the empirical values of these two coefficients, then reconcile the numerical results (i.e., explain how this result arises).
    b) Use an *F* test to test the hypothesis that the three macroeconomic price indexes, *PN*, *PD*, *PS* (remember to take logs) do not have a significant influence on log*GasQPC*.
    c) Use an *F* test to test the hypothesis that the coefficient on the log of the price of gasoline changed in 1974 while all other coefficients in the model were unchanged. Retain the three macroeconomic price indices in your equation.