Quantitative Method

Assignment 9

Due January 17, 2007

1. The following sample moments were computed from 100 observations.

$$\begin{array}{c} \\ \boldsymbol{X'X} = \end{array} \begin{array}{c} \begin{array}{cccc} \textbf{one} & x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 100 & 123 & 96 & 109 \\ 123 & 252 & 125 & 189 \\ 96 & 125 & 167 & 146 \\ 109 & 189 & 146 & 168 \end{bmatrix} \end{array}, \qquad \boldsymbol{X'Y} = \begin{bmatrix} 460 \\ 810 \\ 615 \\ 712 \end{bmatrix},$$

$$(\boldsymbol{X'X})^{-1} = \begin{bmatrix} 0.03767 & & & \\ -0.06263 & 1.129 & & \\ -0.06247 & 1.107 & 1.110 & \\ 0.1003 & -2.192 & -2.170 & 4.292 \end{bmatrix}, \qquad \boldsymbol{Y'Y} = 3924.$$

(a) Compute the ordinary least squares coefficients in the regression of y on a constant, x_1, x_2, and x_3. Construct the ANOVA table. For $\alpha = 0.05$, test (1) $\mathrm{H}_0 : \beta_2 = 0$; (2) $\mathbf{H_0} : \beta_1 + \beta_2 + \beta_3 = 1$ where β_i is the coefficient of x_i.

(b) Compute the ordinary least squares coefficients in the regression of y on a constant, x_1, and x_2, on a constant, x_1, and x_3, and on a constant, x_2, and x_3.

(c) Computer the variance inflation factor (VIF) associated with each variable. If the regressors are collinear, which is the problem variable?

2. Data on the quantity demand of beer (q), the price of beer (P_B), the price of other liquor (P_L), the price of all other remaining goods and services (P_R), and income (m) are given in Table8_3.xls. Consider the log-log demand curve

$$\ln(q) = \beta_1 + \beta_2 \ln(P_B) + \beta_3 \ln(P_L) + \beta_4 \ln(P_R) + \beta_5 \ln(m)$$

(a) Compute the coefficients of the demand relation and construct the ANOVA table.

(b) Use the variance inflation factor (VIF) to examine the presence of multicollinearity.

(c) Construct a 95% ***prediction*** interval for q when $P_B = 3.00$, $P_L = 10$, $P_R = 2.00$ and $m = 50{,}000$. (Hine: construct the interval for $\ln(q)$ and then take anti-logs.)

(d) Construct a 95% ***mean*** interval for q when $P_B = 3.00$, $P_L = 10$, $P_R = 2.00$ and $m = 50{,}000$. (Hint: construct the interval for $\ln(q)$ and then take antilogs.)

(e) Test $H_0 : \beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$.

(f) Test $H_0 : \beta_3 = \beta_4 = 0$.

3. The file *foodus.xls* contains observations on food expenditure y_t, income x_t, and number of persons in each household n_t from a random sample of 38 households in a large U.S. city. Food expenditure and income are measured in terms of thousands of dollars. Consider the statistical model

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 n_t + e_t$$

where the e_t are independent normal random errors with zero mean.

(a) Estimate the above equation using least squares. Report and comment on the results.

(b) Plot the least squares residuals from part 1 against (i) income x_t, and (ii) number of persons n_t. Do these plots suggest anything about the existence of heteroskedasticity?

(c) Use a Goldfeld-Quandt test to test for heteroskedasticity with the observations ordered according to decreasing values of x_t (Split the sample in half such that each has 19 observations). Comment on the outcomes.

(d) Find generalized least squares estimates under the assumption that

$$\mathrm{var}(e_t) = \sigma_t^2 = \sigma^2 \exp\{0.055 x_t + 0.12 n_t\}.$$

Compare the estimates with those obtained using least squares. Does allowing fro heteroskedasticity appear to have improved the precision of estimation?