New York University
Applied Statistics and Econometrics I
Fall 2015

Problem Set 4

Cristian Fuenzalida
cristian.fuenzalida@nyu.edu

(I) The file *DataExample.xls* contains information about the regressors of the following model:

$$\log(G/\text{Pop}) = \alpha + \beta_1 \log(P_g) + \beta_2 \log(\text{Income}/\text{Pop}) + \gamma_1 \log(P_{nc})$$
$$+\gamma_2 \log(P_{uc}) + \tau \cdot \text{month} + \delta_1 \log(P_d) + \epsilon$$

In this exercise you are going to generate artificial data and construct procedures to get an econometric estimation.

1. Generate a sequence of errors with normal standard distribution. Use this sequence and the coefficients in the second sheet to generate an artificial sequence for the dependent variable. Calculate the OLS estimator, the estimated residuals, the estimated variance of the residuals, the asymptotic variance of the OLS estimator, the t-statistics and the standard deviation of each coefficient. Present your results in a single table.
2. Generate now 100 sequences of errors standard normally distributed. We are trying to do repeated samples here. For each of the 100 sequences, generate the implied dependent variable. Reestimate the OLS for each case. Compute the sample variance of the OLS estimator across samples. Plot a histogram of the values of each coefficient. Does this distribution look normal? Should it look normal? Include in each histogram a normal curve with the same mean and variance than the coefficient.
3. Generate now 100 sequences of errors $\epsilon_t \sim \sigma(e_t - 1)$, where e_t is a standard exponential variable with mean 1. Again, we are doing repeated samples here. Repeat the calculations in part 2. Does the distribution look normal in this case? Compare and comment.

(II) The file *datacrime.txt* contains information to estimate a model on county level crime rates. You have to estimate a model using the year 1987 only.

1. Using logarithms of all variables, estimate a model relating the crime rate to the deterrent variables *prbarr*, *prbconv*, *prbpris* and *avgsen*. Include all the t-statistics and the standard errors of the coefficients.
2. Add $log(crmrte)$ for 1986 as an additional explanatory variable, and comment how the elasticities differ from the previous equation and why.
3. Compute the F statistic for joint significance of all of the wage variables (again in logs), using the restricted model from part 2.

(III) The file *invest.txt* contains data on 565 US firms extracted from Compustat for the year 1987. These variables (in order) are: investment to capital ratio (I_i), total market value to asset ratio (Tobin's Q, Q_i), cash flow to asset ratio (C_i) and long term debt to asset ratio (D_i). The flow variables are annual sums for 1987. The stock variables correspond to the beginning of the year.

1. Estimate a linear regression of I_i on the other variables. Calculate appropriate standard errors.
2. Calculate asymptotic confidence intervals for the coefficients.
3. This regression is related to Tobin's Q theory of investment, which suggests that investment should be predicted solely by Q_i. Thus, the coefficient on Q_i should be positive and the others should be zero. Test the joint hypothesis that the coefficients C_i and D_i are zero. Test the hypothesis that the coefficient on Q_i is zero. Are the results consistent with the predictions of the theory?
4. Regress I_i on Q_i, C_i, D_i, Q_i^2, C_i^2, D_i^2, Q_iC_i, Q_iD_i, C_iD_i. Test the joint hypothesis that the six interaction and quadratic coefficients are zero.

(IV) Consider a linear model $y_i = x_i\beta + \epsilon_i$ with normal errors. Let $\theta := (\beta, \sigma^2)'$ and $\hat{\theta}$ the corresponding unrestricted ML estimate. Denote by $\tilde{\theta} := (\tilde{\beta}, \tilde{\sigma}^2)'$ the restricted ML estimate subject to the constraint $R\beta = c$, and:

$$\hat{\Sigma} = \begin{pmatrix} \frac{1}{n\hat{\sigma}^2}\sum_i x_i x_i' & 0 \\ 0 & \frac{1}{2(\hat{\sigma}^2)^2} \end{pmatrix}, \tilde{\Sigma} = \begin{pmatrix} \frac{1}{n\tilde{\sigma}^2}\sum_i x_i x_i' & 0 \\ 0 & \frac{1}{2(\tilde{\sigma}^2)^2} \end{pmatrix}$$

1. Show that $\hat{\beta}$ is just the OLS estimator and $\tilde{\beta}$ is the restricted OLS estimator (that is, $\tilde{\beta}$ minimizes the sum of squared residuals subject to $R\beta = c$.
2. Let $Q_n = \frac{1}{n}\sum_i \ln(f(y_i|x_i, \theta))$. Prove that:
$$Q_n(\hat{\theta}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2} - \frac{1}{2}\ln(SSR_u/n)$$
$$Q_n(\tilde{\theta}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2} - \frac{1}{2}\ln(SSR_r/n)$$
3. Show that the Wald, LM and LR statistics can be written as:
$$W = n\frac{(R\hat{\beta} - c)'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - c)}{SSR_u}$$
$$LM = n\frac{(y - X\tilde{\beta})P_x(y - X\tilde{\beta})}{SSR_r}$$
$$LR = n\left(\ln\left(\frac{SSR_r}{n}\right) - \ln\left(\frac{SSR_u}{n}\right)\right)$$

(V) As we saw in problem IV, in the restricted least squares model the sum of squared residuals is minimized subject to the constraint implied by the null hypothesis $R\beta = r$. Form the Lagrangian as:

$$\mathcal{L} = \frac{1}{2}(y - X\tilde{\beta})'(y - X\tilde{\beta}) + \lambda'(R\tilde{\beta} - r)$$

where λ here is the r-dimensional vector of Lagrange multipliers. Let $\hat{\beta}$ be the restricted estimator of β, the solution of the constrained minimization problem.

- Let b be the unrestricted OLS estimator. Show:

$$\hat{\beta} = b - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(Bb - r)$$

$$\lambda = \left(R(X'X)^{-1}R'\right)^{-1}(Rb - r)$$

- Let $\hat{\epsilon} = y - X\hat{\beta}$, the residuals form the restricted regression. Show that:

$$SSR_R - SSR_U = (b - \hat{\beta})'(X'X)(b - \hat{\beta})$$

$$\begin{aligned} &= (Rb - r)'\left(R(X'X)^{-1}R'\right)^{-1}(Rb - r) \\ &= \lambda' R(X'X)^{-1}R'\lambda \\ &= \hat{\epsilon}' P \hat{\epsilon} \end{aligned}$$

where P is the projection matrix.

- Prove that:

$$\frac{(Rb - r)'\left(R[s^2(X'X)^{-1}]R'\right)^{-1}(Rb - r)}{J} = \frac{(SSR_R - SSR_U)/J}{SSR_U/(n - K)}$$