UNIVERSITY OF ESSEX
Session: 2011/12

DEPARTMENT OF ECONOMICS
João Santos Silva

EC501 Econometric Methods and Applications

Problem Set 2

The Classical Linear Regression Model (continued)

1. Consider the model $y_i = \beta_1 + \beta_2 x_{i2} + \epsilon_i$. Find the OLS estimator of β_1 and β_2 for a dataset where $\mathbf{X}$ and $\mathbf{y}$ are

$$\mathbf{X} = \begin{bmatrix} 1 & 3 \\ 1 & -4 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 4 \\ 1 \\ 2 \\ 2 \end{bmatrix}.$$

2. Let $\mathbf{X}$ be an $n \times k$ matrix of regressors, let $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ and let $\mathbf{M} = \mathbf{I} - \mathbf{P}$.

   (a) Show that $\mathbf{P}$ and $\mathbf{M}$ are symmetric idempotent matrices.

   (b) Find expressions for $\mathbf{PM}$, $\mathbf{MP}$, $\mathbf{M}'\mathbf{P}$, $\mathbf{P}'\mathbf{M}$, $\mathbf{PX}$ and $\mathbf{MX}$.

3. In the linear regression model

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \quad E(\epsilon|\mathbf{X}) = 0, \quad E(\epsilon\epsilon'|\mathbf{X}) = \sigma^2 I_n,$$

   let $\mathbf{b}$ denote the OLS estimator of β and let $\mathbf{e} = \mathbf{y} - \mathbf{Xb}$ denote the vector of estimated residuals. Show that $cov(\mathbf{e}, \mathbf{b}) = E\{\mathbf{e}[\mathbf{b} - E(\mathbf{b})]'\} = 0$ $(n \times K)$.

4. Consider the linear regression model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ where $\mathbf{X}$ is a $n \times K$ matrix, $E(\epsilon|\mathbf{X}) = 0$ and $E(\epsilon\epsilon'|\mathbf{X}) = \sigma^2 I_n$.

   (a) Find $E(\hat{\mathbf{y}}|\mathbf{X})$ and $var(\hat{\mathbf{y}}|\mathbf{X})$, where $\hat{\mathbf{y}} = \mathbf{Xb}$ and $\mathbf{b}$ is the OLS estimator of β.

   (b) The OLS estimator $\mathbf{b}$ has been computed using only the first $n-1$ observations form a sample of size n, and it is desired to forecast the value of the variable $\mathbf{y}$ for observation n using $\hat{y}_n = \mathbf{x}_n'\mathbf{b}$. Find the mean and the variance of the forecast *error*.

P.T.O.

5. A model is estimated by OLS and the residuals $\mathbf{e}_i = \mathbf{y}_i - \mathbf{x}_i'\mathbf{b}$ $(i = 1, \ldots, n)$ are obtained, where $\mathbf{x}_i$ is a $K \times 1$ vector of regressors. Let $RSS = \sum_{i=1}^{n} e_i^2$ denote the residual sum of squares and let $TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$ denote the total sum of squares, where $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$ denotes the sample mean of the dependent variable. Two commonly used measures of fit are

$$R^2 = 1 - \frac{RSS}{TSS} \quad \text{and} \quad \bar{R}^2 = 1 - \frac{RSS/(n-K)}{TSS/(n-1)}.$$

(a) Explain why $0 \leq R^2 \leq 1$.

(b) Show that R^2 and $\bar{R}^2$ are related by the formula

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\frac{(n-1)}{(n-K)}.$$

(c) Suppose that the original model is augmented by a further set of regressors z_i, of dimension $H \times 1$, and is estimated by OLS. Derive a condition (in terms of the respective RSS values) under which the value of $\bar{R}^2$ in the augmented model will be larger than the value of $\bar{R}^2$ in the original model.