WRITTEN PAPER 1 (ECON 4135)

September 13, 2006

Deadline for handling in: September 27, 2006

Consider the following regression model

$$y_i = \alpha + \beta x_i + u_i, \tag{1}$$

where y_i is the dependent variable, x_i is a regressor and u_i is the (unobserved) error term. Furthermore α and β are unknown regression coefficients (parameters) to be estimated.

Assume that the conditional expectation, $E(u_i|x_i)$, and conditional variance, $\text{var}(u_i|x_i)$, are given by

$$E(u_i|x_i) = 0 \text{ and } \text{var}(u_i|x_i) = \sigma_u^2. \tag{2}$$

We have data (y_i, x_i) for $i = 1, ..., n$ individuals. The OLS estimators of β and α are, respectively,

$$\begin{aligned} \widehat{\beta} &= \frac{\sum_i (y_i - \overline{y})(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2} \\ \widehat{\alpha} &= \overline{y} - \widehat{\beta}\,\overline{x}, \end{aligned}$$

where a bar over a variable denotes the empirical mean, e.g. $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

1. What is $E(y_i|x_i)$? Show that (2) implies that $E(u_i) = 0$, $\text{cov}(x_i, u_i) = 0$ and $\text{var}(u_i) = \sigma_u^2$ (Hint: $\text{var}(u_i) = E(\text{var}(u_i|x_i)) + \text{var}(E(u_i|x_i))$)

2. What does it mean that $\widehat{\beta}$ is an unbiased estimator of β? Explain why we can write

$$\widehat{\beta} = \frac{\sum_i (x_i - \overline{x}) y_i}{\sum_i (x_i - \overline{x}) x_i}. \tag{3}$$

Use the expression (3) to show that $E(\widehat{\beta}|X) = \beta$, where $X = (x_1, ..., x_n)$ (i.e. $E(\widehat{\beta}|X)$ is the conditional expectation of $\widehat{\beta}$ given the value of all the regressors, $X = (x_1, ..., x_n)$). Is $\widehat{\beta}$ unbiased? Is $\widehat{\alpha}$ is an unbiased estimator of α?

3. Insert the expression (1) into (3) to show that

$$\widehat{\beta} - \beta = \frac{\sum_i (x_i - \overline{x}) u_i}{\sum_i (x_i - \overline{x}) x_i} \tag{4}$$

Use this to show that, if the u_i are independent across individuals, then

$$\text{var}(\widehat{\beta}|X) = \frac{\sigma_u^2}{\sum_i (x_i - \overline{x})^2}$$

What is the difference in the interpretation of the conditional variance, $\text{var}(\widehat{\beta}|X)$, and the unconditional variance, $\text{var}(\widehat{\beta})$? By the rule of double expectation, $\text{var}(\widehat{\beta}) = E(\text{var}(\widehat{\beta}|X)) + \text{var}(E(\widehat{\beta}|X))$. What is the value of the last expression, $\text{var}(E(\widehat{\beta}|X))$? Can you say something about the properties of $\text{var}(\widehat{\beta}|X)$, as an estimator of $\text{var}(\widehat{\beta})$?

4. What does it mean that $\widehat{\beta}$ is a consistent estimator of β? One can show that $\widehat{\beta}$ is consistent if $\text{var}(\widehat{\beta})$ goes towards 0 when the number of observations, n, increases towards infinity. Show that $\widehat{\beta}$ is consistent if $\sum_{i=1}^{n} (x_i - \overline{x})^2$ goes towards infinity.

5. What does it mean that $\widehat{\beta}$ has an asymptotic normal distribution? What assumptions do we need for this to be the case? Try to use these assumptions to sketch a proof of asymptotic normality of $\widehat{\beta}$ (Hint: use the central limit theorem).

6. Let $\widehat{y}_i$ denote the predicted y_i:

$$\widehat{y}_i = \widehat{\alpha} + \widehat{\beta} x_i.$$

Thus

$$y_i = \widehat{y}_i + \widehat{u}_i,$$

where $\widehat{u}_i$ is called the residual of the regression. Show the following relation:

$$\underbrace{\sum_i (y_i - \overline{y})^2}_{TSS} = \underbrace{\sum_i (\widehat{y}_i - \overline{y})^2}_{ESS} + \underbrace{\sum_i \widehat{u}_i^2}_{SSR},$$

i.e.

$$TSS = ESS + SSR$$

What is the definition of R^2? How can we use R^2 to assess the goodness of fit of a regression equation? Explain intuitively why

$$\widehat{\sigma}_u^2 = \frac{SSR}{n-2}$$

is a meaningful estimator of σ_u^2? What about the estimator $\widetilde{\sigma}_u^2 = \frac{SSR}{n}$? Do you think it matters much in practice whether you use $\widehat{\sigma}_u^2$ or $\widetilde{\sigma}_u^2$?

7. Assume that in a data set we get $\widehat{\beta} = 5$ and $\sqrt{var(\widehat{\beta})} = 3$. Calculate a 95% confidence interval for $\widehat{\beta}$. Set up a test statistic and carry out the following tests:

$$\begin{aligned} i)\ H_0 &: \ \beta = 0 \text{ vs } H_1\text{: } \beta \neq 0 \\ ii)\ H_0 &: \ \beta = 0 \text{ vs } H_1\text{: } \beta < 0 \\ iii)\ H_0 &: \ \beta = 0 \text{ vs } H_1\text{: } \beta > 0 \end{aligned}$$

In which cases do you reject H_0 at the 5% level of significance? Calculate the p-value in each case. What is, in general, the interpretation of a p-value?

8. Now, let us consider a modification of the assumptions (2). Assume instead that

$$E(u_i|x_i) = 0, \text{ var}(u_i|x_i) = x_i^2\sigma_u^2 \tag{5}$$

In this case the error term, u_i, has a variance that depends on x_i. This is called *heteroscedasticity*. Show that the OLS estimator is still unbiased (Hint: Take the expectation of both sides in (4), conditional on X). What is $var(\widehat{\beta}|X)$ in this case (still assuming the u_i are independent)?