

Problem Set 1, Principles of Estimation and Inference in Econometrics  
November 2005

1. Consider the classical linear regression model

$$Y = X\beta + u$$

- i. State clearly the assumptions, and their purpose, underlying the model.
- ii. Derive the OLS estimator for  $\beta$ .
- iii. Show that this OLS estimate is unbiased.
- iv. Derive the variance of this OLS estimator.
- v. State and prove the Gauss-Markov theorem.

2. Consider the following regression model:

$$Y = X\beta + Z\alpha + u$$

- i. Show that, in general, the estimate for  $\beta$  from regressing  $Y$  on  $X$  and  $Z$  is not the same as the estimate of regressing  $Y$  on  $X$  alone.
- ii. Derive the term which captures the bias in the estimated  $\beta$  from the latter approach above.
- iii. Show that regressing the residuals from regressing  $Y$  on  $X$  against the residuals from regressing  $Z$  on  $X$ , produces the same estimate for  $\alpha$  as regressing  $Y$  on  $X$  and  $Z$ .

3. Define the term multicollinearity. What are its symptoms and how can we try to adjust for its presence?

4. Consider the following statement: "Overspecifying the model (i.e., including irrelevant variables) is far less serious than underspecifying the model (i.e., excluding relevant variables) since the latter leads to biased estimates while the former does not." What is your reaction to this comment? Provide algebraic evidence regarding the computation of the estimates and their variances under both approaches.

5. Consider the least squares regression of  $y$  on  $k$  variables (with a constant)  $X$ . Consider an alternative set of regressors  $Z = XP$ , where  $P$  is a nonsingular matrix. Thus, each column of  $Z$  is a mixture of some of the columns of  $X$ . Prove that the residual vectors in the regressions of  $y$  on  $X$  and  $y$  on  $Z$  are identical. What relevance does this have to the question of changing the fit of a regression by changing the units of measurement of the independent variable? (Greene, p. 39).

6. Suppose the 935 males in the data set NLS80 were employees of a single firm that is the defendant in a racial bias law suit. Build the best case possible for the defense. Are you personally convinced by this case?