Econometrics 710
Midterm Exam
March 22, 2007

1. The observations are (y_i, x_{1i}, x_{2i}), $i = 1, ..., n$. You estimate two LS regressions.

$$\begin{aligned} y_i &= x_{1i}'\tilde{\beta}_1 + \tilde{e}_i \\ y_i &= x_{1i}'\hat{\beta}_1 + x_{2i}'\hat{\beta}_2 + \hat{e}_i \end{aligned}$$

and calculate the residual variance estimates

$$\begin{aligned} \tilde{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} \tilde{e}_i^2 \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2. \end{aligned}$$

Show that for any $w \in (0,1)$, there is a constant $a \in (0,1)$ such that

$$\frac{1}{n}\sum_{i=1}^{n} \left(w\hat{e}_i + (1-w)\,\tilde{e}_i\right)^2 = (1-a)\,\hat{\sigma}^2 + a\tilde{\sigma}^2.$$

(Find this constant a.)

Hint: You will need to use the properties of projection matrices.

2. In section 3.8 of the lecture notes, it was shown that if

$$\begin{aligned} y &= X\beta + e \\ E(e \mid X) &= 0 \\ E(ee' \mid X) &= D = \text{diag}\left\{\sigma_1^2, ..., \sigma_n^2\right\} \end{aligned}$$

then

$$E\left(\hat{\sigma}^2 \mid X\right) = \frac{1}{n}\text{tr}\,(MD) \tag{1}$$

where $\hat{\sigma}^2$ is the error variance estimator and $M = I - X\,(X'X)^{-1}\,X'$. Without assuming homoskedasticity, simplify (1) to show that

$$E\left(\hat{\sigma}^2 \mid X\right) = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 - \frac{1}{n}b_n$$

where b_n satisifes $b_n \xrightarrow{p} \text{tr}\left(Q^{-1}\Omega\right)$, where $Q = E\left(x_i x_i'\right)$ and $\Omega = E\left(x_i x_i' e_i^2\right)$.

3. In the model

$$\begin{aligned} y_i &= x_{1i}'\beta_1 + x_{2i}'\beta_2 + e_i \\ E(x_i e_i) &= 0 \end{aligned}$$

where $x_i = (x_{1i}' \; x_{2i}')'$, describe how you would test the hypothesis $H_0 : \beta_1 = \beta_2$ against $H_1 : \beta_1 \neq \beta_2$.

4. Suppose a researcher wants to know which of a set of 20 regressors has an effect on test scores. He regresses test scores on the 20 regressors and reports the results. One of the 20 regressors (study time) has a large t-ratio (about 2.5), while other t-ratios are insignificant (smaller than 2 in absolute value). He argues that the data show that study time is the key predictor for test scores. Do you agree with this conclusion? Is there a deficiency in his reasoning?