Intro to Econometric Theory
Heinz School, Carnegie Mellon University
90-906, Spring 2004-5

Homework #3 (counts as 2 homeworks), due Thursday, April 28

1. Consider the generalized regression model. Suppose we try to run GLS using a matrix Ω_1 which we believe is equal to (or a good estimator of) Ω. Suppose further that we are wrong, and that $\Omega_1 \neq \Omega$. Let's call the estimator we calculate using Ω_1, $\hat{\beta}_{WGLS}$ (W for wrong).

   (a) Is $\hat{\beta}_{WGLS}$ unbiased?

   (b) What is its variance?

   (c) Can you think of a reasonable estimator for this variance?

   (d) Is this variance bigger or smaller than the GLS variance?

   (e) What should we conclude about the dangers of trying to correct for a non-scalar covariance matrix, but screwing it up?

2. With respect to the hospital cost dataset from the web site, please estimate the following two models:

$$lnC = \beta_1 + \beta_2 lnD + \beta_3 lnV + \beta_4 time + \epsilon \quad (1)$$
$$C = \alpha_1 + \alpha_2 D + \alpha_3 V + \alpha_4 time + \epsilon \quad (2)$$

   In the above, D is inpatient days, V is outpatient visits, and *time* is the number of quarters since the first quarter of 1991.

   (a) Discuss how the estimates that these two equations provide for marginal costs differ.

   (b) Test for non-linearity in equation 2. Discuss the meaning of your findings.

   (c) Think carefully about the variable C. Think carefully about equation 2 (not the estimated version, but the theoretical version). Are there circumstances in which equation 2 can give obviously wrong predictions for$C$? Does equation 1 suffer from a similar problem?

   (d) Calculate predicted values for C from both models. Calculate the mean squared error for both models: $\sum(C - \hat{C})$. Which model is better by this criterion?

(e) Would it make sense to compare the R^2 from the two equations in order to decide which is best? Why or why not?

(f) If we care about getting the right answer, which equation should we use and why? (You may do more analysis if you like to answer this question)

3. The Medical Expenditure Panel Survey is an annual survey which collects information about medical expenditures, income, employment, demographics, health information, &c for a representative sample of Americans.

   I have prepared an extract of these data for 1996, and it is available on the course website. The following are the columns in the data, in order:

Variable	Meaning
age	age of person in years
sex	sex of person, 1=male & 0=female
income	income in 1996 $
employed	1=employed, 0=not employed
insured	1=had health insurance, 0=not
health	perceived health status, higher is sicker
spending	spending on health care, 1996 $

   To begin with, let's consider this model:

$$\begin{aligned} \text{spending}_i &= \beta_1 + \beta_2 \text{income} + \beta_3 \text{age} + \beta_4 \text{sex} \\ &+ \beta_4 \text{employed} + \beta_5 \text{insured} + \beta_6 \text{health} \end{aligned} \tag{3}$$

   (a) Test this model for heteroskedasticity. Discuss your findings and and what they mean for inference.

   (b) How much more do people in very good health status spend than do people in excellent health status (estimate and CI). Make sure your answer is correct in the presence of heteroskedasticity.

   (c) Test whether insurance affects spending for people of different health statuses differently and discuss. Make sure your anser is correct in the presence of heteroskedasticity.

4. Do problems 5 and 6 on pp 248-9 of Greene.