Econometrics I
Final Exam.
2003.1.17

1. (25%) State with brief reasons whether the following statements are true, false, or uncertain.

   (a) (5%) In multiple regression, a high correlation in the sample among the regressors (multicollinearity) implies that the least squares estimators of the coefficients are inefficient.

   (b) (5%) Heteroscedasticity in the errors leads to biased estimates of the regression coefficients and their standard errors.

   (c) (5%) Least squares techniques when applied to economic time-series data usually yield biased estimates because many economic time series are autocorrelated.

   (d) (5%) The Durbin-Watson d test assumes that variance of the error term u_t is homoscedastic.

   (e) (5%) You will not obtain a high R^2 value in a multiple regression if all the partial slope coefficients are *individually* statistically insignificant on the basis of the usual t test.

2. (25%) It is known that aggregating data by groups will cause heteroscedasticity. Let the model be

$$Y_{gi} = \beta_1 + \beta_2 X_{gi} + u_{gi}, \; g = 1, \cdots G, \; i = 1, \cdots, n_g$$

   where G is number of groups and n_g is number of observations for group g. The error term u_{gi} is homoscedastic with variance $\text{Var}(u_{gi}) = \sigma^2, \forall g, i$.

   Suppose we aggregate the data by groups and have observations on the average of the variables. The model we can use is,

$$\bar{Y}_g = \beta_1 + \beta_2 \bar{X}_g + \bar{u}_g, \; g = 1, \cdots, G \tag{1}$$

   where $\bar{Y}_g = \frac{\sum_{i=1}^{ng} Y_{gi}}{n_g}$, $\bar{X}_g = \frac{\sum_{i=1}^{ng} X_{gi}}{n_g}$ and $\bar{u}_g = \frac{\sum_{i=1}^{ng} u_{gi}}{n_g}$.

   (a) (5%) Explain why is model (1) heteroscedastic?

   (b) (5%) One way to solve the heteroscedasticity problem and obtain the gnenralized least squares (GLS) estimator is to multiply all the variables by a factor k and make the new error term homoscedastic, then run an OLS regression. What is the factor k we can use such that the transfromed error term $u^* = k\bar{u}_g$ is homoscedastic.

(c) (5%) Describe all the transformed explanatory variables in (b).

(d) (10%) Another interpretation of the GLS estimator in (b) is the **weighted least squares** estimator which weight the square of the residuals for each observation by w_g in the least squares estimation. In other words, we minimize $\sum_{g=1}^{G} w_g \hat{\bar{u}}_g^2$ to obtain the efficient estimators of β_1 and β_2. What is the weight w_g to use?

3. (25%) Consider the following regression-through-the origin model:

$$Y_i = \beta X_i + u_i, \ i = 1, 2$$

You are told that $u_1 \sim N(0, \sigma^2)$ and $u_2 \sim N(0, 2\sigma^2)$ and that they are statistically independent. If $X_1 = +1$ and $X_2 = -1$,

(a) (10%) Obtain the *weighted* lease squares (WLS) estimate of β, $\hat{\beta}_{WLS}$, and its mean and variance. (i.e. minimize the weighted sum of squared residuals.)

(b) (10%) Obtain the OLS estimator of β, $\hat{\beta}_{OLS}$, and its mean and variance.

(c) (5%) Is $\hat{\beta}_{WLS}$ more efficient than $\hat{\beta}_{OLS}$? (**hint:** compare their variances.)

4. (25%) The **von Neumann ratio test** defined as follows can detect autocorrelation. Assuming that the residuals $\hat{u}_t$ are random drawings from normal distribution, von Neumann has shown that for *large* n, the von Neumann ratio

$$vN = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2/(n-1)}{\sum_{t=2}^{n} (\hat{u}_t - \bar{\hat{u}})^2/n}$$

is approximately normally distributed with mean and variance

$$\begin{aligned} \text{E}(vN) &= \frac{2n}{n-1} \\ \text{Var}(vN) &= 4n^2 \frac{n-2}{(n+1)(n-1)^3} \end{aligned}$$

where $\hat{u}_t$ is the residual from OLS estimation.

(a) (7%) What is the relationship between the Durbin-Watson d and vN ? (Note: $\bar{\hat{u}} = 0$ in OLS.)

(b) (8%) It is known that the d statistic lies between 0 and 4. What are the corresponding limits for the von Nemann ratio, vN, when n is sufficiently large?

(c) (10%) Suppose in an application vN was found to be 2.88 with 100 observations. Can we reject the null hypothesis that there is no serial correlation in the data? If yes, is there positive **or** negative serial correlation in the data?