HOMEWORK 1
Fall 2008
Environmental Econometrics
due on Nov. 11 (Tuesday)

Question 1 (Ordinary Least Squares (OLS))
Consider the following simple regression model: for $i = 1, ..., N$,

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

1 Under which assumptions is the ordinary least squares estimator

(a) an unbiased estimator?

(b) the unbiased estimator attaining the minimum variance among the class of unbiased estimators?

2 Derive the exact formula for the OLS estimators of β_0 and β_1.

3 Suppose that each u_i follows independently and identically the normal distribution with zero mean and unknown variance, σ^2. We want to test the following null hypothesis:

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0.$$

Explain a test statistic you will use and what distribution it follows.

4 Suppose that the true regression model is

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + v_i,$$

where $E(v_i|X, Z) = 0$. But you decide to run the above simple regression model between Y and only X (including a constant term). Is the OLS estimator of β_1 unbiased? Discuss in detail.

Question 2 (Hypothesis Testing)
You first regress the log of willingness to pay for a better waste management scheme on the log of income, education levels (low, medium and high) as well as age. And then you regress the log of willingness to pay for a better waste management scheme on the log of income and age, without the dummy variables of education levels. The OLS estimation results are in the table below:

Variables	Coefficient	Std. Err.	Coefficient	Std. Err.
log income	0.17	0.07	0.28	0.02
education low	0.78	0.53		
education medium	0.92	0.56		
education high	1.01	0.59		
age	-0.001	0.004	-0.0008	0.004
R^2	0.049		0.0421	
Number of observations	352			

1 In each regression what is the effect of income on willingness to pay? Is it statistically different from zero? And is it statistically different from 0.3?

2 Discuss the effect of education on the willingness to pay for a better waste management scheme. How much more is the highest education group willing to pay compared to the lowest group?

3 Test the joint significance of the education dummies (i.e., whether they are all zero). You need to set up a null and an alternative hypotheses and derive a test statistic and what distribution it is under what conditions.

Question 3 (Generalized Linear Regression)

You are interested in the relation between variables Y and X and model their relationship with the following regression equation:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \,.$$

A colleague tells you that, given the data, there is a concern about heteroskedasticity.

1 Explain what heteroskedasticity means and the consequences of using OLS.

2 Suppose you use the OLS estimation to obtain an estimate of β_1. Given the concern about heteroskedasticity, you want to get heteroskedasticity-robust standard error of the OLS estimate, $\widehat{\beta}_1$. Please explain how to get it.

3 Suppose you happen to believe that $Var\,(u_i|X)$ is a linear function of X_i such as $Var\,(u_i|X) = \delta_0 + \delta_1 X_i$. Explain how you can test for heteroskedasticity in this case. (Note that you need to specify what the null hypothesis is, what the test statistic is, and its distribution under some assumptions.)

4 Explain if there is an alternative way of implementing an estimation of the parameters, instead of using OLS.

Question 4 (Autcorrelation)

Consider the following simple regression model: for $t = 1, ..., T$,

$$Y_t = \beta_0 + \beta_1 X_t + u_t \,.$$

One colleague tells you that the error term may be autocorrelated.

1 Explain what the autocorrelation means. What are the consequences for the OLS estimation if the error term is indeed autocorrelated?

2 Explain how you would test for it, using the Durbin-Watson test. Under which assumption(s) is (are) this test valid?

3 You are told that u_t follows a moving average process of order 1. Write down the corresponding model for u_t.

4 Compute the covariances between: a) u_t and u_t; b) u_t and u_{t-1}; and c) u_t and u_{t-2}.

Question 5

Krueger (1993 QJE, "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989") examined whether workers who use a computer at work earn a higher wage than otherwise similar workers who do not use a computer at work, using US Current Population Survey data in 1984 and 1989. One of OLS regression equations that he used is

$$\ln W_i = X_i \cdot \beta + \gamma C_i + \varepsilon_i,$$

where W_i denotes hourly wage, X_i denotes a vector of observed characteristics, and ε_i an error for observation $i = 1, ..., N$. C_i represents a dummy variable that equals one if the ith individual uses a computer at work, and zero otherwise. The OLS estimation with the data in 1989 (13,379 people) is shown in the table below, where the estimate of the parameter for C (uses computer at work) is only reported.

	OLS	
Variables	Coefficient	Std. Error
Uses computer at work (1 = yes)	0.188	0.008
R^2	0.451	

1 In 1989 how much did individuals who used a computer at work earn more than those who did not use a computer at work? Is the estimated result statistically significant at the 5% level?

2 One critical concern in interpreting the above OLS regression results is that workers who use computers on the job may be abler workers, and therefore may have earned higher wages even in the absence of computer technology. Suppose that some important variable that is positively correlated with both computer use and hourly wage is omitted. Show why this can result in a bias in the OLS estimation. (For the sake of illustration, you can use a simple regression model, $\ln W_i = \alpha + \gamma C_i + \varepsilon_i$)

3 Explain what R^2 means. Comment on the one reported at the bottom of the table.

Motivated by this concern, Krueger used the following regression equation below:

$$\ln W = X \cdot \beta + \gamma_1 C_w + \gamma_2 C_h + \gamma_3 C_h C_w + \varepsilon,$$

where C_w is a dummy variable that equals one if a worker uses a computer at work and zero otherwise, C_h is a dummy variable that equals one if a worker uses a computer at home and zero otherwise, and $C_h C_w$ is an interaction term between computer use at home and at work. The OLS estimation results with 1989 data is reported below:

	OLS	
Variables	Coefficient	Std. Error
Uses computer at work	0.177	0.009
Uses computer at home	0.070	0.019
Uses computer at home and work	0.017	0.023

4 How much did individuals who used a computer *at work* earn more than those who did not use a computer at all? How much did individuals who used a computer *at home* earn more than those who did not use a computer at all? Are they both statistically significant?

5 How much did individuals who used a computer *at home and at work* earn more than those who used a computer *at work only*?