HOMEWORK 2
Fall 2008
Environmental Econometrics
due on Dec. 2 (Tuesday)

Question 1 (Measurement Error)
Consider the following simple regression equation: for $i = 1, ..., N$,

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

1 Suppose that Y is measured with error. That is, $\widetilde{Y}$ is the observable measure of Y. Let $e = \widetilde{Y} - Y$ denote the measurement error which is assumed to be not correlated with an independent variable X. If you regress observed $\widetilde{Y}$ on X, is the OLS estimator of β_1 biased? Prove your answer.

2 Now, suppose that Y is observed but X is measured with error. That is, $\widetilde{X}$ is the observable measure of X and let $v = \widetilde{X} - X$ denote the measurement error. We assume that $E(v) = 0$ and $Cov(X, v) = 0$. If you regress Y on observed $\widetilde{X}$, is the OLS estimator of β_1 biased? Prove your answer.

3 When the measurement error results in a bias in the OLS estimation, you can use an instrumental variable (IV) to avoid such a bias. What are the definitive properties of IV?

4 Derive an IV estimator for the slope coefficient, β_1.

Question 2 (Simultaneous Equations Model)
You are interested in studying the following hypothesis that more "open" countries have lower inflation rates. In order to test this, you measure the opennes of a country in terms of the average share of imports in gross domestic product, which is denoted by Op. You also notice that the degree of openness might depend on the average inflation rate partly due to governmental policies. Thus, you first consider the following system of two equations:

$$\begin{aligned} Inf_i &= \beta_0 + \beta_1 Op_i + u_i \\ Op_i &= \gamma_0 + \gamma_1 Inf_i + v_i, \end{aligned}$$

where $Cov(u_i, v_i) = 0$ and Inf denotes an annual inflation rate.

1 Derive the reduced form of the model (i.e., express Op_i and Inf_i solely as a function of the coefficients and the error terms).

2 Using the reduced form model, explain why the OLS results in a biased estimator if we regress only the first equation without considering the second equation.

3 Can you identify the model parameters from the above system of the two equations? Now you decide to include two (exogenous) variables into the system: Pc (per capita income) and Ld (the land area of the country in square miles):

$$Inf_i = \beta_0 + \beta_1 Op_i + \beta_2 \log Pc_i + u_i$$
$$Op_i = \gamma_0 + \gamma_1 Inf_i + \gamma_2 \log Pc_i + \gamma_3 \log Ld_i + v_i.$$

Explain which parameters can be identified. Answer these questions using the order condition.

4 You estimated the first equation ($Inf_i = \beta_0 + \beta_1 Op_i + \beta_2 \log Pc_i + u_i$) using the OLS. The OLS estimate was $\widehat{\beta}_1^{OLS} = -0.215\,(0.095)$, where the number in the parentheses is a standard error. Then, you proceeded to estimate the same equation using $\log Ld$ as an instrumental variable for Op. The IV estimate was $\widehat{\beta}_1^{IV} = -0.337\,(0.144)$. Now you wonder whether the difference between the OLS and IV estimates are statistically different. How you can test this?

Question 3

Although schooling and earnings are highly correlated, economists have argued for decades over the causal effect of education on earnings. A careful analysis on this topic requires an exogenous source of variation in education outcomes. Card (1995, "Using Geographic Variation in College Proximity to Estimate the Return to Schooling") used college proximity as an exogenous determinant of schooling.

1 To conduct an investigation of the returns to schooling, he reported the results below estimated by OLS and the method of Instrumental Variable (IV). Each estimation was conducted with other explanatory variables such as experiences and family background, whose results are omitted here. The basic regression equation is as follows:

$$\log(hourly\ wage)_i = \beta_0 + \beta_1 X_{1i} + \ldots + u_i,$$

where X_{1i} is an individual i's years of schooling. The model is estimated in two ways, by OLS and by instrumental variables (IV) using as an instrument for X_{1i} a dummy variable that records whether the individual grew up near a four-year college (called "proximity").

	OLS		IV	
Variables	Coefficient	Std. Error	Coefficient	Std. Error
Education	0.073	0.004		
Proximity			0.132	0.055
R^2	0.300		0.238	

Using the OLS equation, by what percentage will hourly wages increase when there is one year increase of schooling? Is the estimated coefficient statistically significant at the 5% level?

2 Consider the OLS regression equation between log hourly wages and years of schooling. Suppose that some unobservable factor like "ability", which is not included in the equation, affects both years of schooling and log hourly wages. What does this imply the relation between years of schooling and the error term?

3 Under the effect of such an unobservable factor like ability, can you expect that the OLS estimator of β_1 obtained by regressing log hourly wages on years of schooling and other explanatory variables, is unbiased? If not, show why it is not. (For the sake of illustration, you can use a simple regression model, $\log(hourly\ wage)_i = \beta_0 + \beta_1 X_i + u_i$)

4 Provide a formal definition of an instrumental variable (IV). Discuss whether the dummy variable called *proximity* here can be an instrument variable to education.

5 Now, instead of the story of ability, consider the possibility that there is measurement error in years of schooling (In fact, the literature says that 10% of the variance in measured years of schooling is due to measurement error). If there is indeed so, what does it imply between measured years of schooling and the error term? Provide a simple proof of your claim.

6 In the case of measurment error in an independent variable, we have learned the OLS estimator has an attenuation bias. What is an attenuation bias? And what relation does this imply between the OLS estimate and the IV estimate?