MSc in Environmental Economics:
Examination in Environmental Econometrics (GR03)

May 2008

Section 1 is compulsory. Three questions should be chosen from section 2. Section 1 has a weight of 40. Section 2 has a weight of 60, equally divided between each question. Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. Calculators are permitted.

In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the student's first set of answers up to the required number will be the ones that count (not the best answers). All remaining answers will be ignored.

Time: 2:30 hours

Section 1 (Compulsory) [40 points]

Question 1

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 only the estimate of the parameter for C (uses computer at work) is reported.

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

TURN OVER

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*?

Question 2

You are hired by a city government to investigate people's choices of dwelling between three types: house (H), apartment (A), and low-cost flat (F). For each $k = H, A, F$, a latent random utility is determined by

$$U_k = \beta_{k0} + \beta_{k1} Age + \beta_{k2} Gender + \beta_{k3} \log income + u_k.$$

CONTINUED

Suppose that you choose the choice of low-cost flat (F) as the baseline and decide to use a multinomial logit model.

1 What are the probabilities of choice H and choice A?

2 Write down the log odds-ratio of choice H (House) with respect to choice F (low-cost flat) as a function of the parameters of the model and the explanatory variables. Does the log odds-ratio between H and F depend on the parameters of choice A (Apartment)?

3 The estimation results you obtained are reported below:

Choice of House			Choice of Apartment		
	Coeff.	Std. Err.		Coeff.	Std. Err.
age	0.027	0.010	age	0.002	0.012
gender	-0.409	0.259	gender	-0.305	0.297
log income	1.358	0.186	log income	1.495	0.216
constant	-10.753	1.560	constant	-11.703	1.820

What is the marginal effect of income on the choice of House (using the log odds-ratio)? Compute the change of the log odds-ratio when log income increases from 1 to 1.5.

4 One colleague claims that you should not use a multinomial logit model but use an ordered probit model. What is your response? (Note that you must first explain the differences between logit model and ordered probit model and then present your opinion.)

Section 2 [60 points]

Note: Choose three questions out of four questions.

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

TURN OVER

2 Explain how you would test for autocorrelation, 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 2 (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 3 (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 openness 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).

CONTINUED

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.

2 Answer the following questions using the order condition. 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):

$$\begin{aligned} Inf_i &= \beta_0 + \beta_1 Op_i + \beta_2 \log Pc_i + u_i \\ Op_i &= \gamma_0 + \gamma_1 Inf_i + \gamma_2 \log Pc_i + \gamma_3 \log Ld_i + v_i. \end{aligned}$$

Explain which parameters can be identified.

3 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 4 (Hypothesis Testing)
Consider the follwing regression equation:

$$\ln Hprice_i = \beta_0 + \beta_1 \ln Nox_i + \beta_2 Room_i + \beta_3 Stratio_i + \beta_4 \ln Dist_i + u_i,$$

where $Hprice$ denotes a median housing price in a community, Nox the amount of nitrogen oxide in the air, $Room$ the average number of rooms in houses, $Stratio$ the average student-teacher ratio of schools, and $Dist$ a weighted distance of the community from employment centers. The estimation results are in the table below (the second regression has only $\ln Nox$ as an explanatory variable):

Variables	Coeff.	Std. Err.	Coeff.	Std. Err.
$\ln Nox$	-0.954	0.117	-1.532	0.230
$Room$	0.255	0.019		
$Stratio$	-0.052	0.006		
$\ln Dist$	-0.134	0.043		
Constant	11.08	0.32	9.077	0.432
R^2	0.581		0.402	
Number of observations	352			

TURN OVER

1 In the full regression equation, what is the effect of nitrogen oxide on housing price? Is it statistically significantly different from zero? from -1?

2 Discuss the effect of *Room* (*Stratio* and *Dist*, respectively) on housing price.

3 Test the joint significance of *Room*, *Stratio*, and *Dist* variables (i.e., whether they are all zero). (See the appendix for some critical values)

4 Under what assumptions are these tests valid? Suppose the estimation is unbiased. What could jeopardize your conclusions?

Appendix: Some useful test statistics

5% Confidence Levels	
$t_{\alpha/2}$	1.96
$F(2, 348)$	3.00
$F(3, 347)$	2.60
$F(4, 347)$	2.37

END OF PAPER