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

May 2007

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 questions. Fully justified answers are required to obtain high marks. However, answers are not expected to exceed 10-15 lines. Calculators are permitted.

Time: 2:30 hours

Section 1 (Compulsory) [40 points]

Question 1

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.

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:

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

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	

TURN OVER

1 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? [5 points]

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? [5 points]

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, $\log(\textit{hourly wage})_i = \beta_0+\beta_1 X_i+u_i$) [5 points]

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 points]

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. [5 points]

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? [5 points]

Question 2

Consider the following system of two equations:

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

where $Cov\,(u_i, v_i) = 0$.

1 From this structural model, derive the reduced form of the model. 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. [5 points]

CONTINUED

2 Can we identify the model parameters from the above system of two equations? Now consider we have an exogenous variable in the second equation as follows:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + u_i \\ X_i &= \gamma_0 + \gamma_1 Y_i + \gamma_2 Z_i + v_i. \end{aligned}$$

Explain which equation can be identified. Answer these questions using the order condition.[5 points]

Section 2 [60 points]

Note: Choose three questions out of four questions.

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? [2.5 points]

(b) the unbiased estimator attaining the minimum variance among the class of unbiased estimators? [2.5 points]

2 Derive the exact formula for the OLS estimators of β_0 and β_1. [5 points]

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. [5 points]

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. [5 points]

TURN OVER

Question 2 (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. [5 points]

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. [5 points]

3 Suppose you happen to believe that $Var\left(u_i|X\right)$ is a linear function of X_i such as $Var\left(u_i|X\right) = \delta_0 + \delta_1 X_i$. 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.) [5 points]

4 Explain if there is an alternative way of implementing an estimation of the parameters, instead of using OLS. [5 points]

Question 3 (Binary Response Model)

You have a survey data set about environmental concerns in a community. The data set contains two variables: Y_i is a Yes ($Y_i = 1$)/No ($Y_i = 0$) response on the question "Are you concerned about the environment in your community?", X_{1i} is an individual i's log of annual income and X_{2i} is an individual i's age. Suppose you consider the following regression equation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \, .$$

1 Suppose that using OLS, you get an estimate of the slope coefficient, $\widehat{\beta}_1 = 0.1$.

(a) Explain how to interpret the result. [2.5 points]

(b) Explain possible problems of using OLS. [2.5 points]

2 Now you decide to use the logit model. What is the probability of responding Yes given X_{1i} and X_{2i} in the logit model? And derive the log odds-ratio between Yes and No. [5 points]

CONTINUED

3 Suppose the table below displays the maximum likelihood estimation results from the logit model.

Variable	Coefficient	Standard errors
Log income (X_1)	0.48	0.12
Age (X_2)	0.03	0.01
Constant	-5.07	1.16
Log likelihood	- 278.84	

(a) How do you interpret the coefficient on log income? Is it significantly different from zero? [2.5 points]

(b) What is the log odds-ratio for an individual with log income equal to 5 and age equal to 40? Explain what it means. [2.5 points]

4 Now you want to test a hypothesis that the environmental concern does not depend on income level and age.

(a) Specify the null and alternative hypotheses. [2.5 points]

(b) Suppose you got the following maximum likelihood estimation by omitting the two variables, log income and age, from the logit regression equation.

Variable	Coefficient	Standard Errors
Constant	- 3.96	0.98
Log likelihood	- 286.85	

Specify your test statistic to test the null hypothesis and explain what your conclusion is. [2.5 points]

Question 4 (Ordered Probit Model)

You want to evaluate the determinant of the attitudes toward the environment. You collect information on individuals living in a community who report their environmental concerns as either 'not concerned at all', 'somewhat concerned', or 'very concerned' (coded as $0, 1, 2$, respectively). Explanatory variables are log of income, age, sex and atmospheric pollution (odour or "smell"). The results are tabulated in the table below.

Variables	Coefficient	Standard Error
log income	0.274	0.080
age	0.021	0.005
sex	0.023	0.125
smell	0.363	0.138
cut I	0.096	0.746
cut II	2.984	0.697

TURN OVER

1 About the model.

(a) Write down the model which has been estimated. [4 points]

(b) Explain what the variables *cut I* and *cut II* refer to. [2 points]

(c) How would you write the probability of being "very concerned"? [3 points]

2 What is the sign of the marginal effect of log income on the probabilities of each outcome? (Note that the marginal effect of X_1 (log income) on each probability is given by

$$\begin{aligned}
\frac{\partial \Pr(Y=0)}{\partial X_1} &= -\beta_1 \phi(\mu_1 - \beta X) \\
\frac{\partial \Pr(Y=1)}{\partial X_1} &= \beta_1 \left[\phi(\mu_1 - \beta X) - \phi(\mu_2 - \beta X)\right] \\
\frac{\partial \Pr(Y=2)}{\partial X_1} &= \beta_1 \phi(\mu_2 - \beta X),
\end{aligned}$$

where βX represents $\beta_1 \log income + \beta_2 sex + \beta_3 age + \beta_4 smell$. How do you interpret your answers? [6 points]

3 A colleague of yours tells you that you should run a multinomial estimation, instead of the ordered probit model. Present your opinion. [5 points]

Appendix: Some useful test statistics

5% Confidence Levels	
$t_{\alpha/2}$	1.96
$\mathcal{X}^2_\alpha(2)$	5.99
$\mathcal{X}^2_\alpha(10)$	18.31

END OF PAPER