Exam #2

Economics 435

Spring 2002

Please make certain I can understand your answer. For example, circle or put boxes around your final answers, be careful to write clearly, etc. If I don't understand you, I will assume you are wrong.

Also, be sure to ask me if you are uncertain over what a question means.

1 School spending

A famous review article by Eric Hanushek (Journal of Economic Literature, 1986), looked at the results from a large number of studies that estimated the effect of increased spending on student performance. The models estimated took the form

$$score = \beta_0 + \beta_1 * spending + \beta_2 * x + u$$

where *score* is the average score in the school on some standardized test, *spending* is per-pupil spending in dollars, and x is some other variable (or collection of variables).

Hanushek looked at about 25 studies of this form, all of which used slightly different data sets and approaches. In 20 studies, the estimate of β_1 was statistically insignificant. In 4 of the studies, the estimate was positive and significant. In one study the estimate was negative and significant.

a) Suppose that spending can be considered exogenous. Can we conclude from these 25 studies that higher spending does not improve student performance? Explain.

b) Of course, spending is not exogenous. Identify one reason why spending may be positively correlated with u, and one reason why it may be negatively correlated with u.

2 Data mining

Suppose that a researcher has our cross-country growth data set and decides to engage in a little of what is called "data mining." His data set consists of information on each country in the 98-country sample. He wants to know what country characteristics are associated with economic growth, and decides to do the following. Let y be the average annual growth rate of per capita real GDP, over the period 1960-1985. For each other variable x in the data set he estimates a regression of the form:

$$y = \beta_0 + \beta_1 x + u$$

If he finds a significant (at the 5% level of significance) coefficient on x, he writes up the results in a scholarly article and publishes the article in an economics journal.[1] If he does not find a significant

[1] Any article, no matter how bad, can be published in *some* journal.

coefficient on x, he does nothing. Suppose that there are 40 potential explanatory variables in his data set, so he estimates 40 separate regressions and potentially publishes 40 separate articles.

Without the researcher's knowledge, his research assistant has accidentally replaced each x with a completely random number.

a) One of the variables is the number of economists per capita. What is the probability that the researcher will publish an article claiming to find a *negative* relationship between the number of economists per capita and economic growth?

b) What is the expected number of articles published by the researcher?

c) What is the probability that the researcher will not publish any articles?

3 Private school

In a recent article, Evans and Schwab (1995) studied the erffect of attending a private high school on the probability of attending university. Let *univ* be a binary variable equil to one if a student attends university, and zero otherwise. Let *priv* be a binary variable equal to one if the student attends a private high school. They are interested in a linear probability model of the form:

$$univ = \beta_0 + \beta_1 * priv + \beta_2 x + u$$

where x is some other factor or factors.

a) Why might *priv* be correlated with u?

b) Let $Cath$ be a binary variable equal to one if the student is Catholic. Evans and Schwab use $Cath$ as an IV for *priv*. Identify two conditions which must hold in order for $Cath$ to be a valid IV.

c) Which of these two conditions can be tested? How would you do so?

4 A variation on the Hausman test

You are working on an econometrics project with a classmate, who we will call Fred. The project involves a standard IV estimation procedure. You are interested in estimating the model:

$$y = \beta_0 + \beta_1 x + u$$

You have a random sample, and some variation in x, but you are concerned that x is endogenous ($corr(u, x) \neq 0$). You have an instrumental variable z that is exogenous and relevant.

Fred is in charge of running Shazam. In order to estimate β_1, he uses 2SLS. First he uses OLS to estimate the model:

$$x = \pi_0 + \pi_1 z + v$$

Then he calculates $\hat{x} = \hat{\pi}_0 + \hat{\pi}_1 z$, where $\hat{\pi}_0$ and $\hat{\pi}_1$ are the OLS estimates. He then estimates an OLS regression of y on $\hat{x}$ to get the 2SLS estimates of β_0 and β_1.

For your project, you are also supposed to perform a Hausman test for the exogeneity of x. Ordinarily this involves calculating $\hat{v} = x - \hat{x}$, and estimating the OLS regression of y on x and $\hat{v}$. The null hypothesis that x is exogenous is equivalent to the null hypothesis that the coefficient on $\hat{v}$ is zero.

However, Fred has made a potentially disastrous mistake. Instead of doing the standard Hausman test, he estimated the following model by OLS:

$$y = \lambda_0 + \lambda_1 x + \lambda_2 \hat{x} + \epsilon$$

What's worse, he accidentally deleted all of the data right after printing out the results from this regression, and you don't have time to get the data back.

Let $(\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2)$ be the coefficient estimates, and let

$$\begin{bmatrix} \hat{\sigma}_0^2 & \hat{\sigma}_{01} & \hat{\sigma}_{02} \\ \hat{\sigma}_{01} & \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{02} & \hat{\sigma}_{12} & \hat{\sigma}_2^2 \end{bmatrix}$$

be their estimated covariance matrix.

This question will ask you to use these results to devise a hypothesis test which is equivalent to the Hausman test.

a) State the null hypothesis that x is exogenous in terms of parameter values. State the alternative hypothesis as well.

b) In terms of the $\hat{\lambda}$'s and $\hat{\sigma}$'s, define a test statistic that has an asymptotic standard normal distribution under the null hypothesis.

5 Employer-provided health insurance and "job lock"

This question is related to the health insurance market in the United States. Since their system is quite different from Canada's I'll give you a little background.[2]

1. The government does not provide healthcare or health insurance.
2. Many, but not all, employers provide health insurance as part of their compensation package. Usually, health insurance comes with higher-paying jobs – in other words, McDonald's doesn't provide health insurance to workers, but Microsoft does.
3. When an employer provides health insurance to a worker, that insurance is provided to the worker's entire immediate family (spouse and children) if they do not already have their own insurance.
4. Because of adverse selection,[3] it is very expensive for individuals to get health insurance.
5. When a person changes employers, that person must change to the new employer's health insurance companies.

[2] In order to keep the problem as simple as possible while preserving some of the important issues, some of these statements are not exactly correct, or are no longer correct. For the purposes of this problem, pretend that they are.

[3] "Adverse selection" is a fancy name for the following problem with insurance. People often know more about their own health than their insurance companies. As a result, the people with the highest demand for insurance will be people who are already sick, and thus people who impose high costs on their insurers. As a result, insurers have to price their products assuming that their customers are in below-average health. Employers get a better deal than individuals because they sign contracts with insurers to insure all of their employees, so the adverse selection problem is less.

6. Health insurers do not cover "pre-existing conditions." In other words, if you change jobs (and thus insurers), any condition (cancer, HIV, broken leg) that you already had at the start of your new job will not be covered by your new health insurer.

Policymakers are concerned that this regime creates "job-lock" – situations where a worker cannot change jobs because he or she has a pre-existing condition that is covered at his current job, but would not be covered at any new job.

Brigitte Madrian, in an article in the 1992 Quarterly Journal of Economics, used data on employer-provided health insurance to test the job-lock hypothesis. Madrian has a random sample of married male workers who were employed in January 1990, and has three variables for each worker: y is a binary variable indicating whether a worker voluntarily changed jobs in 1990, x is a binary variable indicating whether the worker's January 1990 employer provided health insurance, and z is a variable indicating whether the worker's wife had employer-provided health insurance.

a) One possible way of determining whether the current system creates job-lock is to see if workers in jobs with health insurance are less likely to change jobs than workers in jobs without health is to estimate the following linear probability model using OLS:

$$y = \beta_0 + \beta_1 x + u$$

Unfortunately, it is likely that x is correlated with u. Give one reason why you might expect this to be the case.

b) Suppose that:

$$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz + u$$

where $E(u|x, z) = 0$. Let $\bar{y}_{ij}$ be the average of y for all observations in which $x = i$ and $z = j$ and let n_{ij} be the number of observations in which $x = i$ and $z = j$. For example, n_{00} is the number of workers in the sample that do not have health insurance and whose wives do not have health insurance either, and $\bar{y}_{00}$ is the fraction of such workers that changed jobs.

You wish to estimate the β's, but you only have the following data: $\bar{y}_{00}$, $\bar{y}_{01}$, $\bar{y}_{10}$, $\bar{y}_{11}$, n_{00}, n_{01}, n_{10}, and n_{11}

Construct consistent estimates of β_0, β_1, β_2 and β_3 using these eight numbers (Hint: find the plim of each average, and then solve).

c) This problem involves a lot of algebra and is time consuming, so you might want to leave it for last. Madrian argues that the job-lock hypothesis corresponds to the hypothesis that β_3 is negative. Calculate the variance of your estimate of β_3, using these eight numbers. (Note that the covariance of two sample averages from two different populations is zero).

d) Use your answers to the previous two questions to construct a test statistic that has an asymptotic standard normal distribution when $\beta_3 = 0$.