Exam #1

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 The lazy professor (12 pts)

An econometrics professor is giving an exam to a class of 300 students, and is very lazy. After grading the first 50, he notices a pattern: most of the students who did well on the exam wrote in pen rather than in pencil, and that most who got the first question did well on the exam. He estimates a simple regression:

$$s = \beta_0 + \beta_1 p + \beta_2 f + u \tag{1}$$

where s is the score on the entire exam, p is an indicator variable which is equal to one if the exam is written in pen, and f is the score on the first question.

With the remaining 250 students, he grades the first question, determines whether they wrote with pencil or pen, and calculates their predicted score given (p, f):

$$\hat{s} = \hat{\beta}_0 + \hat{\beta}_1 p + \hat{\beta}_2 f$$

He then enters $\hat{s}$ as the person's score in the gradebook.

The students, naturally, are upset when they find out about this. The professor explains "Exam scores are just an estimate of how well a student understands the material. This is just a different estimate. As long as my estimate is unbiased, what's to complain about?"

a) Even if $\hat{s}$ is an unbiased estimator of how well the student understands the material, why (in statistical terms) would it still be an inferior estimator to grading the entire exam?

b) The professor admits he plans to use the same approach on the second exam. Specifically, he will use the same model and the same coefficient estimates (he will not re-estimate the coefficients), but he will grade the first question on the second exam and assign each person the grade $s_i = \hat{\beta}_0 + \hat{\beta}_1 p_i + \hat{\beta}_2 f_i$ What do you expect to happen?

c) The professor admits he plans to use the same approach on the second exam as on the first exam. Unlike in the previous question, he will grade the entire exam for the first 50 students, use this to get new estimates of $(\beta_0, \beta_1, \beta_2)$, and use those estimates to construct predicted exam scores. Will the professor even be able to estimate these coefficients? Explain.

2 Cigarette taxes (12 pts)

The provincial government announced last Tuesday at noon that the excise tax on cigarettes would be increased by 80 cents per pack, effective at midnight that night. The tobacco industry is very upset about this and has commissioned an econometrician to study the effect on demand.

The econmetrician collected cigarette sales data on both Tuesday (the day of the announcement) and Wednesday (the day after the announcement) from a large random sample of British Columbia stores that sell tobacco. Each store-day combination is a single observation, for example, the Safeway in my neighborhood on Tuesday is one observation and the same Safeway on Wednesday is a different observation. Let c be the number of cigarettes sold, and let w be an indicator variable equalling one on Wednesday. He estimates a regression:

$$\ln c = \beta_0 + \beta_1 w + u$$

He finds that $\hat{\beta}_1$ is negative, statistically significant, and quite large. Specifically, $\hat{\beta}_1 = -0.5$.

a) The tobacco companies report that this study indicates that the new tax will result in a . . . percent decline in their business. Fill in the blank.

b) You are an econometrician working for the provincial government. Explain what is wrong with this study.

c) Propose an improvement to their methodology.

3 Forecasting hydro demand (16 pts)

You are a manager at BC Hydro, and you employ two econometricians to forecast the demand for electricity. Let e be monthly energy sales per household in kilowatt-hours, let p be that day's price of energy in cents per kilowatt-hour, and let t be the temperature at noon. Econometrician #1 estimates the following model:
$$e = \beta_0 + \beta_1 p + \beta_2 t + u$$

while econometrician #2 estimates:

$$\ln e = \gamma_0 + \gamma_1 \ln p + \gamma_2 \ln t + u$$

a) Currently, the typical household consumes 900 kilowatt-hours per month, and the price is 5.7 cents per kilowatt-hour. You are considering a price increase to 6 cents per kilowatt-hour, and ask econometrician #1 for an prediction of the change in energy demand (sales per household in kilowatt-hours) at the new price. Find his answer, in terms of his coefficient estimates $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$.

b) Suppose that you ask econometrician #2 the same question. Give her answer in terms of her coefficient estimates $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$. In order to make the calculation easier, assume that $\hat{\gamma}_1 = -1$.

c) Do you expect to get the exact same answer from both econometricians? Explain.

d) Under what condition on γ_1 will a price increase lead to an increase in revenue?

4   Absolute convergence (28 pts)

A famous 1986 paper by William Baumol investigated the question of whether there was absolute convergence in per capita real GDP among the countries of the world over the period 1870 to 1979. His data set consists of real per capita GDP for 1870 ($RGDP70$) and 1979 ($RGDP79$) for the 16 industrialized countries which have data. He estimates the model:

$$(\ln RGDP79 - \ln RGDP70) = \beta_0 + \beta_1 \ln RGDP70 + u$$

using OLS. His results are:

Coefficient	Estimate	Standard Error
$\hat{\beta}_0$	8.457	1.05
$\hat{\beta}_1$	-0.995	0.394

a)   The absolute convergence hypothesis holds that $\beta_1 < 0$. Perform a one-tailed asymptotic hypothesis test for absolute convergence. Be sure to clearly state your null hypothesis, alternative hypothesis, significance level, test statistic, distribution of test statistic, critical value(s), and test result.

b)   Calculate and report the p-value associated with this test.

c)   Calculate and report a 95% asymptotic confidence interval for β_1.

d)   Another researcher, Robert Barro, points out that countries which were poor in 1870 do not tend to have data on 1870 GDP. Will this lead to results which are biased in favor of convergence, against convergence, or not biased at all? Explain.

e)   A third researcher, Brad DeLong, points out that no one was gathering data on GDP in 1870, so all 1870 GDP data have been gathered recently. As a result, only countries which were relatively rich in 1979 have gone back and gathered data on 1870 GDP. Will this lead to results which are biased in favor of convergence, against convergence, or not biased at all? Explain.

f)   A fourth researcher, Lantt Pritchett, points out that we can actually impute a lower bound on 1870 GDP. The way we do this is to calculate the minimum caloric intake needed to survive and reproduce (while it is possible to live at the edge of starvation for a long time, the female reproductive system shuts down well before the level of starvation that will kill a person). Figure out the cost of this intake, and let that be MINGDP. We know, for example, that 1870 per capita GDP in India in 1870 must be at least MINGDP, because the Indian population has survived.

Pritchett then sets RGDP70 equal to MINGDP for all countries that do not have 1870 real per capita GDP data, and runs the same regression as Baumol. Pritchett finds that the coefficient on RGDP70 is positive and significant. Will Pritchett's procedure lead to results which are biased in favor of convergence, against convergence, or not biased at all? Explain.

g)   Given Pritchett's results, does the evidence support the hypothesis of absolute convergence in the period 1870-1979? Explain.

5   Miscellaneous questions (16 pts)

a) Does multicollinearity lead to biased OLS estimates?

b) You estimate the following regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

You would like to test the null hypothesis that $\beta_1 = 2\beta_2$. Calculate the appropriate asymptotic test statistic as a function of $\hat{\beta}_1, \hat{\beta}_2$, $\widehat{var}(\hat{\beta}_1)$, $\widehat{var}(\hat{\beta}_2)$, and $\widehat{cov}(\hat{\beta}_1, \hat{\beta}_2)$.

c) You flip a coin 3 times. Let x be the number of times the coin comes up "heads". Write down (be careful to get it exactly right) and plot the CDF of x.

6   Teen drinking (16 pts)

On Tuesday, the National Center on Addiction and Substance Abuse (CASA) released a study that indicated that 25 percent of all alcohol consumed in the US is consumed by underaged (below the legal drinking age of 21) drinkers. Here's the first few paragraphs of a newspaper report on the study:

> *WASHINGTON (AP) – Nearly a third of high school students say they binge drink at least once a month, according to a report that says underage drinkers now account for 25 percent of the alcohol consumed in this country.*
>
> *"Underage drinking has reached epidemic proportions in America," said Joseph Califano Jr., president of the National Center on Addiction and Substance Abuse at Columbia University, which issued the report Tuesday.*
>
> *The report, which analyzes two years' research, "is a clarion call for national mobilization to curb underage drinking," said Califano, a former U.S. Secretary of Health, Education and Welfare.*
>
> *Some of the report's findings:*
>
> - *Eighty-seven percent of adults who drink had their first drink before age 21.*
> - *The gender gap for drinking is disappearing. Female ninth-graders were just as likely to be drinkers as male ninth-graders.*
> - *Eighty-one percent of high school students have consumed alcohol, compared with 70 percent who have smoked cigarettes and 47 percent who have used marijuana.*
> - *Most teens who experiment with alcohol continue using it. Among high school seniors who had tried alcohol, 91.3 percent still were drinking in the 12th grade.*
>
> *The percentage of teens who drink on binges – 31 percent among high school students – was obtained by using the Youth Risk Behavior Survey of the Centers for Disease Control and Prevention in Atlanta, published in 2000. The conclusion that underage drinkers accounted for 25 percent of alcohol consumption was based on the 1998 National Household Survey on Drug Abuse. The number of drinks consumed by underage drinkers in a month was divided by the total number of drinks in the same period for the sample.*

Here's a little background on the study. The study is based on the 1998 National Household Survey on Drug Abuse (NHSDA). The NHSDA uses a sample design which is stratified by age and oversamples young people. In particular, the NHSDA has 25,500 respondents over age 12, 9,759 (38%) of whom are between age 12 and 20. (the drinking age in the US is 21, so anyone in this group is "underage"). Based on the 2000 US Census, 19% of the over-12 population is between 12 and 20. NHSDA data is collected by interviewing the respondent at home on various questions related to drug and alcohol use.

Let y be the total number of drinks consumed by an individual in a month, and let x be an indicator variable which equals one if the respondent is between age 12 and 20.

a) Let Y be the total number of drinks per month consumed by respondents in the survey, and let Y^{12-20} be the total number of drinks per month consumed by age 12-20 respondents. As the story indicates, the 25 percent figure was calculated by dividing Y^{12-20} by Y. In other words:

$$\frac{Y^{12-20}}{Y} = \frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} y_i} = 0.25$$

Suppose that the average number of drinks per month in the survey is 10. What is the average number of drinks per month among those aged 12-20?

b) Shortly after this story appeared in newspapers (I remember seeing it on BCTV Tuesday night), liquor industry representatives pointed out a serious methodological flaw in the procedure used to obtain the "25% of alcohol is consumed by underaged" result. Please explain that flaw.

c) Calculate and report the correct estimate.

d) Of the other four findings (the ones in the bulletted list above), which if any would be affected by that same methodological flaw?