Exam #1

Economics 435: Quantitative Methods

Fall 2008

1 A few warmup questions

a) Prove that if $E(u|x) = E(u)$, then $cov(x, u) = 0$.

b) Prove that $cov(x, u) = 0$ does not necessarily imply that $E(u|x) = E(u)$.

c) Find the slope ($\frac{dy}{dx}$) and elasticity ($\frac{dy}{dx}\frac{x}{y}$) in terms of x (i.e., "y" should not appear in your answer) when $\ln y = \beta_0 + \beta_1 x$.

d) For each city in Table 1 below, find the probability of snow in the air on Christmas, given that there is at least 2 cm of snow on the ground.

Location	Probability of a white Christmas	Probability of a perfect Christmas
Quebec City	100%	50%
Vancouver	11%	4%

Table 1: A "white Christmas" is defined as there being at least 2 cm of snow on the ground on December 25 of this year. A "perfect Christmas" is defined as there being snow in the air and at least 2 cm of snow on the ground on December 25 of this year. Source: Environment Canada, `http://www.msc-smc.ec.gc.ca/media/xmas/prob_e.html`.

e) For each city in Table 1, find a lower bound and upper bound on the probability of snow in the air on Christmas. Hint: All probabilities lie between zero and one.

2 The relationship between least squares prediction and the expected value

In answering this question, remember that you can find the minimum of a convex function $f(.)$ by solving the equation $f'(a) = 0$ for a.

a) Let x be a random variable. Let m be defined as the number that minimizes the expected squared prediction error (ESPE), where:

$$ESPE \equiv E\left[(x - m)^2\right]$$

Prove that $m = E(x)$.

b) Now let $D_n = \{x_1, x_2, \ldots, x_n\}$ be a random sample of size n on the random variable x. Let $\hat{m}$ be defined as the number that minimizes the average squared prediction error (ASPE), where:

$$ASPE \equiv \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{m})^2$$

Prove that $\hat{m} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

3 The education production function

Many economic studies of education are aimed at estimating the *education production function*. The idea is that a student's academic achievement - usually measured by a test score - is a function of various costly educational "inputs" such as teacher quality, class size, etc. If we can estimate this function, we might be able to improve student outcomes by allocating funds to those inputs that have the highest impact per dollar on achievement.

Suppose that our model of the education production function is:

$$c = \beta_0 + \beta_1 q + \beta_2 s + u$$

where c is the student's current level of achievement, q is some variable that measures the "quality" of the student's educational environment this school year, and s is the student's level of achievement at the beginning of this school year. This particular model is called the "value-added" model of the educational production function. We assume that:

$$E(u|q, s) = 0$$

and we would like to estimate β_1.

a) Suppose that we only have a random sample of data on (c, q) - i.e., we do not have data on s - and that we estimate an OLS regression of c on q. Let $\hat{\beta}_1^A$ be the estimated regression coefficient on q. Find plim $\hat{\beta}_1^A$.

b) Based on the result above, what direction is the asymptotic bias[1] in $\hat{\beta}_1^A$? Be explicit about the assumptions that you are making, and be sure to justify them.

c) Now suppose we have a random sample of data on (c, q, s). One commonly applied method for estimating the value-added model is to estimate a regression of the "gain score" $g \equiv (c - s)$ on q. Let $\hat{\beta}_1^B$ be the estimated regression coefficient on q. Find plim $\hat{\beta}_1^B$.

d) The "gain score" approach for estimating the educational production function is usually interpreted as imposing the additional assumption that past inputs affect current achievement with no "decay", i.e., that $\beta_2 = 1$. When $\beta_2 = 1$, is $\hat{\beta}_1^B$ a consistent estimator of β_1?

e) When $\beta_2 < 1$, what is the direction of the asymptotic bias in $\hat{\beta}_1^B$? Use the assumptions you made for part (b) of this question.

f) Now suppose instead we have a random sample of data on $(c, q, \tilde{s})$ where $\tilde{s}$ is the student's score on a test given at the beginning of the term. We will interpret this test score as a measure of the true achievement level with classical measurement error, i.e.:

$$\tilde{s} = s + \epsilon$$

where

$$E(\epsilon|q, s, u) = 0$$

We will also assume[2] that:

$$E(s|q, \tilde{s}) = a_0 + a_1 q + a_2 \tilde{s}$$

where

$$a_1 = var(\epsilon)cov(q, s)var(s)var(q)\left(1 - corr(q, s)^2\right)$$

and $0 < a_2 < 1$.

[1] The asymptotic bias of an estimator $\hat{\theta}$ is just plim $(\hat{\theta} - \theta)$.

[2] In case you're wondering, this assumption is actually harmless, as a slightly weaker version of the statement follows directly from the assumption of classical measurement error, and is sufficient to establish the results below.

Suppose that we estimate an OLS regression of c on $(q, \tilde{s})$. Let $\hat{\beta}_1^C$. be the coefficient on q from that regression. Find plim $\hat{\beta}_1^C$.

g) What is the direction of the asymptotic bias in $\hat{\beta}_1^C$? Use the assumptions you made for part (b), along with any other assumptions you need.

h) Now suppose we estimate an OLS regression of $\tilde{g} \equiv c - \tilde{s}$ on q. Find plim $\hat{\beta}_1^D$ (hint: you've already found the answer to this question).

i) Under what condition on $var(\epsilon)$ and β_2 does the gain score approach estimate β_1 with less asymptotic bias? State your answer both in math and in words (as best you can).

j) Suppose that you are sure that $0 < \beta_2 \leq 1$ and $cov(s, q) > 0$, and that you have data on $(c, q, \tilde{s})$. Using the results so far, construct an interval estimate of β_1, i.e., a finite-width interval calculated from the data[3] that will contain β_1 with probability approaching 1 as the sample size approaches infinity.

k) Now suppose that we have data on the true values of (c, q, s). However, some students have missing data on c because they did not take the test. For each of the following cases, identify whether an OLS regression of c on (q, s) is sufficient to consistently estimate β_1.

1. An unexpected flu outbreak kept some students away from school that day.
2. School administrators, under pressure to get high scores, have kept students who performed poorly on the previous exam (i.e., had low values of s) from taking this year's exam.
3. School administrators, under pressure to get high scores, have asked teachers to keep students out of this year's exam whenever they expect (based on experience with the student) them to perform poorly.
4. Leaders in a particular visible minority community organize a boycott, arguing that the test is discriminatory[4] because average scores in this community are below average. The result of the boycott is that almost no students from that community takes the exams.

[3] Two things to note about this question:

- It's hard. Don't be discouraged if you don't know how to answer it.
- The correct answer does not require additional calculation.

[4] For the purpose of this question, please assume that the test is not discriminatory.