Econometrics - Midterm Exam 2001 - Sketch of Answers

Problem 1. (15 points)

There are many different reasons why one might think that the returns to education (in terms of wages) depends on the skill level of the individual. In a recent paper, a labor economist operationalized this by considering the following model for the relationship between wages (denoted by y) and education (denoted by x),

$$y_i = \beta_0 + x_i \beta_i + \varepsilon_i$$

where the coefficient that measures returns to education is modeled as

$$\beta_i = \gamma_0 + x_i \gamma_1 + v_i$$

Here i denotes individual,

$$E[\varepsilon_i | x_i] = 0$$

and

$$E[v_i | x_i] = 0$$

It is clear that this implies that

$$E[y_i | x_i] = \beta_0 + x_i \gamma_0 + x_i^2 \gamma_1$$

Let $\mu = E[x]$ and assume that $E\left[(x_i - \mu)^3\right] = 0$.

Suppose that the model above is correct, but that one mistakenly measures the returns to education by the slope coefficient in the least squares regression of y_i on a constant and x_i. Show that under random sampling, this will yield an estimate of $\gamma_0 + 2\gamma_1 \mu$. (In other words, the slope coefficient in the least squares regression of y_i on a constant and x_i will estimate the average derivative of the conditional mean of y with respect to x.)

(Hint: first show that $E\left[(x_i - \mu)^3\right] = 0$ implies that $E\left[x_i^3 - \mu x_i^2\right] = 2\mu V[x]$)

(more space for Problem 1 on this page)

First note that

$$\begin{aligned} E\left[(x-\mu)^3\right] &= E\left[x^3 - 3\mu x^2 + 3\mu^2 x - \mu^3\right] \\ &= E\left[x^3\right] - 3\mu E\left[x^2\right] + 3\mu^3 - \mu^3 \\ &= E\left[x^3\right] - 3\mu E\left[x^2\right] + 2\mu^3 \\ &= E\left[x^3\right] - \mu E\left[x^2\right] - 2\mu E\left[x^2\right] + 2\mu^3 \\ &= E\left[x^3 - \mu x^2\right] - 2\mu\left(E\left[x^2\right] - \mu^2\right) \\ &= E\left[x^3 - \mu x^2\right] - 2\mu V(x) \end{aligned}$$

so $E\left[(x-\mu)^3\right] = 0$ implies

$$E\left[x^3 - \mu x^2\right] = 2\mu V(x)$$

Second, recall that the slope coefficient in the proposed regression converges to

$$\begin{aligned} \frac{cov(y,x)}{V(x)} &= \frac{E[yx] - E[y]E[x]}{V(x)} \\ &= \frac{\left(\beta_0 E[x] + \gamma_0 E[x^2] + \gamma_1 E[x^3]\right) - \left(\beta_0 + \gamma_0 E[x] + \gamma_1 E[x^2]\right)E[x]}{V(x)} \\ &= \frac{\gamma_0\left(E[x^2] - E[x]^2\right) + \gamma_1\left(E[x^3] - E[x]E[x^2]\right)}{V(x)} \\ &= \gamma_0 \frac{V(x)}{V(x)} + \gamma_1 \frac{2E[x]V(x)}{V(x)} = \gamma_0 + 2\gamma_1 E[x] \end{aligned}$$

Problem 2. (18 points)

The following two models are estimated by OLS

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \varepsilon_i \tag{2-1}$$

$$y_i = \alpha_0 + (x_{1i} + x_{2i})\,\alpha_1 + \varepsilon_i \tag{2-2}$$

The results are summarized

Equation 1.		
Variable	Coefficient	Std. Error
β_0	1.51	0.19
β_1	-0.32	0.66
β_2	2.06	0.64
R-squared	0.34	
Sum squared resid	55.81	
Number of observations	100	

Equation 1.		
Variable	Coefficient	Std. Error
α_0	1.49	0.20
α_1	0.89	0.13
R-squared	0.32	
Sum squared resid	57.84	
Number of observations	100	

Answer the following questions under the assumption that the usual regularity conditions for OLS are satisfies.

(a) (3 points) Consider Equation 1. Construct a 95% confidence interval for β_2

CI is

$$2.06 \pm 1.96 \cdot 0.64 = (0.81, 3.31)$$

if the (asymptotic) normal distribution is used.

It is

$$2.06 \pm 1.98 \cdot 0.64 = (0.79, 3.33)$$

if one uses the $t-$distribution

(b) (3 points) Test whether β_1 in Equation 1 equals 0 (against the alternative that it differs from 0). Test at a 5% level of significance.

$$T = -\frac{0.32}{0.66} = -0.48$$

(not significant)

(c) (6 points) Test whether $\beta_1 = \beta_2$ in Equation 1 (test at a 5% level of significance).

Note that equation 2-2 is the restricted regression. The test–statistic is

$$F = \frac{(57.84 - 55.81)/1}{55.81/(100-3)} = 3.5282$$

(not significant — the critical value is around 3.84)

You could also use the reported R^2. That would in principle give the same result. But since the numbers are rounded, the actual result would differ.

(d) (8 points) Do you have enough information to construct a 95% confidence interval for $\beta_1 - \beta_2$? If not, explain why. Otherwise, construct the confidence interval.

The F-stat above is the square of the T-stat for testing $\beta_1 - \beta_2 = 0$. The can also be written as

$$T = \frac{\widehat{\beta}_1 - \widehat{\beta}_2}{se\left(\widehat{\beta}_1 - \widehat{\beta}_2\right)}$$

Since $\widehat{\beta}_1 - \widehat{\beta}_2 = 2.38$, this implies that

$$se\left(\widehat{\beta}_1 - \widehat{\beta}_2\right) = \frac{2.38}{\sqrt{3.5282}} = 1.27$$

So the 95% CI is

$$2.38 \pm 1.96 \cdot 1.27$$

Problem 3. (15 points)

Let y be a random variable and let x be a random vector of dimension K. Let β be a K dimensional vector of constants, and define $\varepsilon = y - x'\beta$. Assume that $E[x_i \varepsilon_i]$ and $E[x_i]$ are finite.

Suppose that there are n draws from the distribution of (y, x) denoted by (y_i, x_i), $i = 1, \ldots n$, and that the stochastic process (y_i, x_i) is jointly stationary and ergodic.

If $\widetilde{\beta}$ is *any* consistent estimator for β, is it true that $\frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i'\widetilde{\beta}\right)^2 \xrightarrow{p} E[\varepsilon_i^2]$? (Prove or give a counter example).

(HINT: Only 2 points will be subtracted if you answer the question under the stronger assumption that (y_i, x_i) is i.i.d. rather than "jointly stationary and ergodic"

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i'\widehat{\beta}\right)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left((y_i - x_i'\beta) - x_i'\left(\widehat{\beta}-\beta\right)\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i - x_i'\left(\widehat{\beta}-\beta\right)\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 2\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i x_i'\right)\left(\widehat{\beta}-\beta\right) + \left(\widehat{\beta}-\beta\right)'\left(\frac{1}{n}\sum_{i=1}^{n}x_i x_i'\right)\left(\widehat{\beta}-\beta\right)
\end{aligned}
$$

If $E[x_i x_i']$ is finite, then the ansswer is "yes" as the assumptions allow one to apply law of large numbers to all the averages. (This answer alone will give full credit)

If $E[x_i x_i']$ is not finite, then the answer is "no" as $\frac{1}{n}\sum_{i=1}^{n} x_i x_i'$ will then diverge, and a $\widehat{\beta}$ for which $\left(\widehat{\beta}-\beta\right)$ converges to 0 slowly enough will cause the last terms to not converge to 0.

Problem 4. (12 points)

Consider the regression

$$y_t = \beta_1 + y_{t-1}\beta_2 + \varepsilon_t \tag{4-1}$$

where

$$\varepsilon_t = \alpha u_{t-1} + u_t$$

where the u_t's are a sequence of independent and identically distributed random variables with $E[u_t] = 0$ and $V[u_t] = 1$. Assume that $-1 < \alpha < 1$.

(a) (6 points) Suppose you regress y_t on y_{t-1} and a constant. Discuss the properties of the OLS estimator in this case. Be as explicit as you can.

It is inconsistent as y_{t-1} is correlated with ε_t

(b) (6 points) Discuss how your answer to (a) would change if (1) was replaced by

$$y_t = \beta_1 + y_{t-2}\beta_2 + \varepsilon_t$$

and you regressed y_t on y_{t-2} and a constant.

y_{t-2} is uncorrelated with ε_t, so this regression will yield a consistent estimator. But it will not be efficient and standard results will not imply that it will be unbiased.

(25) 5. A researcher has 10 i.i.d. draws from a $N(0,\sigma^2)$ distribution, $(X_1, X_2, \ldots, X_{10})$. He is interested in the competing hypotheses:

$$H_0: \sigma^2 = 1 \text{ vs. } H_A: \sigma^2 = 4$$

He plans to test these hypotheses using the test statistic $\hat{\sigma}^2 = \frac{1}{10}\sum_{i=1}^{10} X_i^2$. The procedure is to reject H_0 in favor of H_A if $\hat{\sigma}^2 > 1.83$; otherwise H_0 will not be rejected.

(7.5) (a) Determine the size of this test. (If you can't do the calculation exactly, tell me exactly how you would determine the size.)

IF Ho is TRUE

$$\Sigma X_i^2 \sim \chi^2_{10}$$

Rejection Region is $\frac{1}{10}\Sigma X_i^2 > 1.83$ or $\Sigma X_i^2 > 18.3$

$$P(\chi^2_{10} > 18.3) = .05$$

(7.5) (b) Determine the power of this test. (If you can't do the calculation exactly, tell me exactly how you would determine the power.)

IF alternative is correct then

$$\sum \left(\frac{X_i}{2}\right)^2 \sim \chi^2_{10} \quad \text{or} \quad \frac{1}{4} \sum X_i^2 \sim \chi^2_{10}$$

Now $\frac{1}{10} \sum X_i^2 > 1.83 \Rightarrow \sum X_i^2 > 18.3 \Rightarrow$

$$\frac{1}{4} \sum X_i^2 > \frac{18.3}{4} = 4.57$$

$$P(\chi^2_{10} > 4.57) = .92$$

(10) (c) Is the researcher using the most powerful test? If so, prove it. If not, construct a more powerful test.

$$L(\sigma^2 | X) = \text{CONSTANT} - \frac{1}{2} \ln \sigma^2 - \frac{1}{2} \frac{1}{\sigma^2} \sum X_i^2$$

$\Rightarrow$

$$L(4 | X) = \quad '' \quad - \frac{1}{2} \ln 4 - \frac{1}{2} \frac{1}{4} \sum X_i^2$$

$$L(1 | X) = \quad '' \quad - \frac{1}{2} \ln 2 - \frac{1}{2} \sum X_i^2$$

LR Test: Reject when $L(4|X) - L(1|X)$ is Large

This mean

$$-\frac{1}{2} \frac{1}{4} \sum X_i^2 + \frac{1}{2} \sum X_i^2 + \text{CONSTANT} \quad \text{is Large}$$

or $\sum X_i^2$ is Large.

Thus LR critical Region is $\sum X_i^2 >$ CRITICAL VALUE.

This is the same test that is used by the Researcher.

(13) 6. Suppose that X_i is i.i.d. with mean μ, variance σ^2 and finite fourth moment. A random sample of 100 observations are collected, and you are given the following summary statistics:

$\bar{X} = 1.23$ and $s^2 = 4.33$

(where $s^2 = (N-1)^{-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$ is the sample variance.)

(8) (a) Construct an approximate 95% confidence interval for μ. Why is this an "approximate" confidence interval?

$\bar{X} \overset{A}{\sim} N(\mu, \frac{1}{n}s^2)$

95% CI is $1.23 \pm 1.96\left(\frac{4.33}{100}\right)^{\frac{1}{2}}$

or $1.23 \pm .41$ or $.82$ to 1.64

IT is approximate because it uses the Normal approximation for the dist of $\bar{X}$, which is based on the CLT

(5) (b) Let $\alpha = \mu^2$. Construct an approximate 95% confidence interval for α.

Let $\hat{\alpha} = \bar{X}^2 = 1.23^2$

Then $\hat{\alpha} \overset{A}{\sim} N\left(\alpha, \left(\frac{d\alpha}{d\mu}\right)^2 \cdot VAR(\bar{X})\right)$

or

$\hat{\alpha} \overset{A}{\sim} N\left(\alpha, (2\mu)^2 \cdot \frac{1}{n}\sigma^2\right)$ or

$\hat{\alpha} \overset{A}{\sim} N\left(\alpha, (2\bar{X})^2 \frac{4.33}{100}\right)$ or $\hat{\alpha} \sim N(\alpha, .26)$

95% CI is

$\hat{\alpha} \pm 1.96(.26)^{\frac{1}{2}}$ or $1.51 \pm 1.96(.51)$

1.51 ± 1 or $.51$ to 2.51

(This is based on the δ-method -- other methods can also be used.)

(12) 7. (6) (a) Suppose that $(1-0.8L)y_t = \varepsilon_t$, where ε_t is white noise with unit variance. Derive the Wold representation for y.

$$y_t = (1-.8L)^{-1}\varepsilon_t$$
$$= \varepsilon_t + .8\,\varepsilon_{t-1} + .8^2\varepsilon_{t-2} + .8^3\varepsilon_{t-3} + \dots,$$

(6) (b) Suppose that $y_t = (1-3L)\,\varepsilon_t$, where ε_t is white noise with unit variance. Derive the Wold representation for y.

Invertible Representation is $y_t = (1-\frac{1}{3}L)\,a_t$

$a_t \sim$ white Noise with

σ_a^2 solves $\quad (1+(\frac{1}{3})^2)\,\sigma_a^2 = (1+3^2)\,\sigma_\varepsilon^2$

$$\Rightarrow \sigma_a^2 = 9\,\sigma_\varepsilon^2 = 9$$

(10) 8. Suppose that X_t follows the MA(2) process

$$X_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

where μ is a constant and ε_t is i.i.d. with $E(\varepsilon_t)=0$, $E(\varepsilon_t^2)=\sigma^2$. Let $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$.

Show that $\bar{X} \xrightarrow{p} \mu$.

$$\bar{X} = \frac{1}{n}\Sigma X_i = \mu + \frac{1}{n}\Sigma(\varepsilon_i + \theta\varepsilon_{i-1})$$

$$= \mu + \frac{1}{n}\varepsilon_H + \frac{1}{n}\sum_{i=1}^{T-1}(1+\theta)\varepsilon_i + \frac{1}{n}\theta\varepsilon_0$$

$\Rightarrow$

$E(\bar{X}) = \mu$

$VAR(\bar{X}) = \frac{1}{n^2}\sigma^2 + (1+\theta)^2\frac{1}{n}\sigma^2 + \frac{1}{n^2}\theta^2\sigma^2 \longrightarrow 0$

$\Rightarrow \bar{X} \xrightarrow{m.s.} \mu \Rightarrow \bar{X} \xrightarrow{p} \mu.$

or, easier

$$\bar{X} - \mu = \underbrace{\frac{1}{n}\Sigma\varepsilon_i}_{\downarrow p\ 0} + \theta\underbrace{\frac{1}{n}\Sigma\varepsilon_{i-1}}_{\downarrow p\ 0} \Rightarrow \bar{X} \xrightarrow{p} \mu$$