Final Exam
Economics 401
Thurs., May 13, 2009

<u>Show All Work</u>. Only partial credit will be given for correct answers if we can not figure out how they were derived.

Points:

Problem 1:	20
Problem 2:	20
Problem 3:	15
Problem 4:	15
Problem 5:	15
Problem 6:	15
Total:	100

Problem 1: Consider the following output from Stata:

```
. reg hrsemp grant lsales lemploy

      Source |       SS       df       MS              Number of obs =     320
-------------+------------------------------           F(  3,   316) =   34.03
       Model |  54669.0521      3  18223.0174          Prob > F      =  0.0000
    Residual |  169231.177    316  535.541699          R-squared     =  0.2442
-------------+------------------------------           Adj R-squared =  0.2370
       Total |  223900.229    319  701.881595          Root MSE      =  23.142


------------------------------------------------------------------------------
      hrsemp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |   34.36007   3.724866     9.22   0.000         ???         ???
      lsales |   1.516521   2.053283     0.74   0.461     -2.523312    5.556354
     lemploy |  -6.655931   2.234891    -2.98   0.003     -11.05308   -2.258783
       _cons |   11.57259   24.86771     0.47   0.642     -37.35462    60.49981
------------------------------------------------------------------------------
```

where you have data on the amount of job training received by employees for firms and whether the firm received a job training grant.

- hrsemp: hours of training per employee at the firm level
- grant: dummy variable for whether the firm received a job training grant
- lsales: the log of annual sales for the firm
- lemploy: the log of the total number of employees.

a) Construct a 95% confidence interval for the coefficient on grant.

b) Interpret this confidence interval in a descriptive manner. That is if we think of this regression as descriptive, what do I learn from this regression?

c) Now interpret the confidence interval as a causal parameter.

d) Do you believe this is a causal parameter, why or why not?

Problem 2: Suppose you are interested in the effect of exercise on weight. Assume the causal model takes the form

$$W_i = \beta_0 + \beta_1 E_i + u_i$$

where W_i is the weight of individual i and E_i is the average amount of time individual i exercises each week.

a) Explain what the parameter β_1 means as a causal parameter.

b) If you simply ran this regression, do you think you would obtain an unbiased estimate? Do you think the bias would be positive or negative? (please explain why)

c) Now suppose we have data on the current unemployment rate in the town in which individual i lives. Explain why this might make sense as a plausible instrumental variable.

d) Calling this variable Z_i explain how you would use it as an instrumental variable. (You don't need to solve explicitly, but give me two equations in the two unknown parameters $\widehat{\beta_0}$ and $\widehat{\beta_1}$.)

e) Is this a perfect instrument? What might be the problem?

Problem 3: Suppose you want to compare the average wages of sociology majors and economics majors.

Let

$$log(w_i^s) = \beta_0^s + \beta_1^s GPA_i + u_i^s$$
$$log(w_i^e) = \beta_0^e + \beta_1^e GPA_i + u_i^e$$

where w_i^s and w_i^e are the wages received by sociology and economics majors respectively, and GPA_i is college GPA. Explain how to test whether these regression functions are the same. That is, explain how to run one regression and use it to test whether the slope and intercept parameters for sociology and economics majors are jointly the same.

Problem 4: You are interested in the regression

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t.$$

Throughout we will maintain the assumptions that this is the right causal model, that there is no multicollinearity, and that

$$E(u_t \mid X_{1it}, X_{2t}) = 0.$$

Consider the following four different models for the error term:

i) u_t is independent and identically distributed.

ii) $u_t = X_{1t}^2 \varepsilon_t$ where ε_t is independent and identically distributed.

iii) $u_t = \rho u_{t-1} + \varepsilon_t$ where ε_t is independent and identically distributed.

iv) $u_t = \rho u_{t-1} + X_{1t}^2 \varepsilon_t$ where ε_t is independent and identically distributed.

Please explain the consequences of OLS in each of these four cases (that is are they unbiased, etc.). For cases in which there is a problem, suggest a better alternative.

Problem 5: Suppose you have a sample $Y_1, ..., Y_N$ and you want to estimate the expected value of Y.

a) Suggest an estimator that is unbiased, but not consistent.

b) Suggest an estimator that is consistent, but not unbiased.

Problem 6: Consider the following system of equations

$$Y_{1i} = \alpha_1 Y_{2i} + \alpha_2 Y_{3i} + \beta_0 + u_{1i}$$
$$Y_{2i} = \alpha_3 Y_{1i} + \gamma_0 + \gamma_1 X_i + u_{2i}$$
$$Y_{3i} = \delta_0 + \delta_1 W_i + u_{3i}$$

where

$$E(u_{1i} \mid X_i, W_i) = 0$$
$$E(u_{2i} \mid X_i, W_i) = 0$$
$$E(u_{3i} \mid X_i, W_i) = 0.$$

Suggest a procedure to obtain consistent estimates for α_1 and α_2.