Elisabeth Sadoulet and Erin Kelley

ENV ECON 118 / IAS 118 - Introductory Applied Econometrics
Assignment 6
Due Thursday December 3 at the beginning of class

Exercise 1 (10 points): Final 2011 #2

Consider the following model of the effect of aid on growth:

$$GDPG_i = \beta_0 + \beta_1 AID_i + \beta_2 GOV_i + \beta_3 GOV_i \times AID_i + u_i$$

where $GDPG_i$ is the growth rate of GDP per capita in country i, AID_i is the amount of public aid received by country i, and GOV_i is an index of quality of governance for country i that varies from 0 for very poor governance to 1 for excellent governance.

1. What is the marginal effect of AID on growth for a country with excellent governance $GOV = 1$?

2. How would you proceed to estimate a confidence interval for this marginal effect?

Exercise 2 (10 points): Final 2014 #5

Does Medicare save lives? Adults 65 years and older are eligible for Medicare health insurance, while those under 65 are not. You have a cross-sectional dataset of emergency room visits made by adults ages 63-67 in California in 2000. This dataset includes the birthday, gender, and family income of the patient, as well as whether or not the patient died within seven days of the emergency room visit.

(a) How would you estimate the causal effect of Medicare on deaths? Be sure to write down the exact regression you would run and define each variable in your regression. [You can use a linear probability model here, for simplicity.] State which coefficient in your regression will give you the estimated causal effect.

(b) What key assumption do you need to make for your regression in part (a) to estimate the causal effect of Medicare on the probability of dying after an emergency room visit?

Exercise 3 (10 points): Final 2013 #10

Following are two logit estimations of school enrollment of children between 10 and 15 years old. The variables are defined as follows:

enroll	=1 if child enrolled in school, 0 otherwise
age	age in years
male	=1 if male, 0 otherwise
distsec	distance to the closest school, in km
headeduc	education of the household head, in years
hhsize	family size

(a) Use these results to test the hypothesis that neither of the two variables headeduc and hhsize affects the probability of enrollment at the 5% significance level.

(b) Using the results of Model B, how does the distance to school affect the probability of school enrollment?

Model A

```
. logit enroll age male distsec headeduc hhsize

Logistic regression                               Number of obs   =       1128
                                                  LR chi2(5)      =     245.80
                                                  Prob > chi2     =     0.0000
Log likelihood = -429.42657                       Pseudo R2       =     0.2225
------------------------------------------------------------------------------
      enroll |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.7785121   .0606438   -12.84   0.000    -.8973719   -.6596524
        male |   .5626309   .1739383     3.23   0.001     .2217182    .9035437
     distsec |  -.1587534     .03403    -4.67   0.000     -.225451   -.0920557
    headeduc |   .0644678   .0395693     1.63   0.103    -.0130866    .1420222
      hhsize |   .0012207   .0381183     0.03   0.974    -.0734898    .0759311
       _cons |   11.44994   .8720377    13.13   0.000      9.74078    13.15911
------------------------------------------------------------------------------
```

Model B

```
. logit enroll age male distsec

Logistic regression                               Number of obs   =       1128
                                                  LR chi2(3)      =     242.95
                                                  Prob > chi2     =     0.0000
Log likelihood = -431.12895                       Pseudo R2       =     0.2194
------------------------------------------------------------------------------
      enroll |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.7796616   .0603568   -12.92   0.000    -.8979587   -.6613645
        male |   .5592701   .1731976     3.23   0.001     .2198091    .8987311
     distsec |  -.1635575   .0338092    -4.84   0.000    -.2298223   -.0972927
       _cons |   11.64867   .8263081    14.10   0.000     10.02913     13.2682
------------------------------------------------------------------------------

. mfx

Marginal effects after logit
      y  = Pr(enroll) (predict)
         =  .87480617
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |  -.0853888      .00599  -14.25   0.000  -.097136 -.073642   12.2972
    male*|   .0617935      .01938    3.19   0.001   .023807   .09978   .510582
 distsec |  -.0179129      .00369   -4.85   0.000  -.025147 -.010678   2.46057
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Exercise 4 (15 points): Final 2009 #6

Using panel data from 22 cities in Indiana over the period from 1981 to 1988, you want to estimate the effect of the enterprise zone program on unemployment. The variable is a dummy variable equal to 1 if the city i has an enterprise zone in year t, the variable is the number of unemployment claims filed during year t in city i, and d81, d82, , d88 are dummy variables for the years 1981 to 1988.

```
Model A

. reg loguclms ez

Model B

. xtreg loguclms ez, i(city) fe

Model C

. xtreg loguclms ez d82-d88, i(city) fe
```

1. Write the equations for the models A-C that would be estimated with these Stata commands [Be very careful with indices].
2. What does model B control for that was a possible source of bias in estimating the causal effect of ez with model A? What does model C control for that was a possible source of bias in estimating the causal effect of ez with model A?
3. From the estimation of Model C, what do you conclude about the effect of the construction of an enterprise zone on unemployment? Explain why it differs from the coefficient estimated in model B [Hint: think about the correlation between the ez variable and time].

```
Model A

. reg luclms ez

      Source |       SS       df       MS              Number of obs =     198
-------------+------------------------------           F(  1,   196) =   16.69
       Model |  7.88619577     1  7.88619577           Prob > F      =  0.0001
    Residual |  92.6100832   196  .472500425           R-squared     =  0.0785
-------------+------------------------------           Adj R-squared =  0.0738
       Total |  100.496279   197  .510133396           Root MSE      =  .68739

------------------------------------------------------------------------------
     loguclms |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ez |  -.4725695   .1156733    -4.09   0.000    -.7006935   -.2444455
       _cons |   11.30057   .0557544   202.68   0.000     11.19062    11.41053
------------------------------------------------------------------------------
```

Model B

```
. xtreg loguclms ez, i(city) fe
Fixed-effects (within) regression               Number of obs      =       198
Group variable (i): city                        Number of groups   =        22
R-sq:  within  = 0.3083                         Obs per group: min =         9
       between = 0.0002                                        avg =       9.0
       overall = 0.0785                                        max =         9
                                                F(1,175)           =     78.00
corr(u_i, Xb)  = -0.2147                        Prob > F           =    0.0000
------------------------------------------------------------------------------
    loguclms |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ez |  -.7668601   .0868293    -8.83   0.000    -.9382276   -.5954927
       _cons |   11.36894   .0353991   321.17   0.000     11.29908    11.43881
-------------+----------------------------------------------------------------
     sigma_u |  .59249639
     sigma_e |  .40931737
         rho |  .67693198   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(21, 175) =    17.99             Prob > F = 0.0000
```

Model C

```
. xtreg loguclms ez d82-d88, i(city) fe
Fixed-effects (within) regression               Number of obs      =       198
Group variable (i): city                        Number of groups   =        22
R-sq:  within  = 0.8148                         Obs per group: min =         9
       between = 0.0002                                        avg =       9.0
       overall = 0.3415                                        max =         9
                                                F(8,168)           =     92.36
corr(u_i, Xb)  = -0.0040                        Prob > F           =    0.0000
------------------------------------------------------------------------------
    loguclms |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ez |  -.1044148    .059753    -1.75   0.082    -.2223782    .0135486
         d82 |   .2963117   .0564519     5.25   0.000     .1848651    .4077582
         d83 |  -.0584394   .0564519    -1.04   0.302     -.169886    .0530071
         d84 |  -.4183358    .058757    -7.12   0.000     -.534333   -.3023386
         d85 |  -.4309709   .0626459    -6.88   0.000    -.5546455   -.3072963
         d86 |  -.4604488   .0626459    -7.35   0.000    -.5841234   -.3367742
         d87 |  -.7281326   .0626459   -11.62   0.000    -.8518072    -.604458
         d88 |  -1.066817   .0626459   -17.03   0.000    -1.190492   -.9431425
       _cons |   11.53358   .0325925   353.87   0.000     11.46923    11.59792
-------------+----------------------------------------------------------------
     sigma_u |  .55551522
     sigma_e |  .21619434
         rho |  .86846297   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(21, 168) =    59.31             Prob > F = 0.0000
```

Exercise 5 (15 points): Final 2012 # 3

Following are the results of two estimations for the wage of college students, where lwage is log hourly wage, *college* is the number of credits completed at college, exper is years of work experience, $black = 1$ if African-American, $hispanic = 1$ if Hispanic, and $white = 1$ if neither African-American or Hispanic.

(a) Test the hypothesis at the 5% level that there is no race effect in wage determination?

(b) How do the wages of Hispanic workers compare to the wages of white workers, and of African-American workers?

(c) How would you set up an equation that will give you the standard error on the difference in predicted wage between Hispanic and African-American workers of same education and experience?

```
. reg lwage college exper

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  2,   197) =   36.61
       Model |  12.9069342     2  6.45346708            Prob > F      =  0.0000
    Residual |  34.7216717   197  .176252141           R-squared     =  0.2710
-------------+------------------------------           Adj R-squared =  0.2636
       Total |  47.6286059   199  .239339728           Root MSE      =  .41982

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     college |   .0906426   .0141812     6.39   0.000     .0626762    .1186091
       exper |   .0055367   .0008727     6.34   0.000     .0038156    .0072578
       _cons |   1.435584   .1148637    12.50   0.000     1.209063    1.662104
------------------------------------------------------------------------------

. reg lwage college exper black hispanic

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  4,   195) =   19.72
       Model |  13.7174678     4  3.42936694            Prob > F      =  0.0000
    Residual |  33.9111382   195  .173903273           R-squared     =  0.2880
-------------+------------------------------           Adj R-squared =  0.2734
       Total |  47.6286059   199  .239339728           Root MSE      =  .41702

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     college |    .087449   .0141936     6.16   0.000     .0594563    .1154416
       exper |   .0051825   .0008823     5.87   0.000     .0034424    .0069226
       black |  -.2038212    .094475    -2.16   0.032    -.3901452   -.0174972
    hispanic |   -.046364   .2434926    -0.19   0.849    -.5265811    .4338531
       _cons |   1.508139   .1190169    12.67   0.000     1.273413    1.742865
------------------------------------------------------------------------------
```

Exercise 6 (15 points): Final 2011 #4

In July 1998, ARCO acquired a major chain of gas stations, inducing a concern that it may have reduced competition and allowed itself to increase prices in the Summer of 1998. Following are average prices collected at Arco and other gas stations (the other gas stations are those not acquired by Arco) in the Los Angeles area. All prices are in $/gallon

	May-98	Oct-98
ARCO gas stations	1.26	1.43
Other gas stations	1.29	1.41

(a) What is the difference-in-differences estimate of the impact of the ARCO acquisition on gasoline prices? Compute the value and interpret it.

(b) What equation would you estimate and what test would you perform to show that the acquisition has produced a statistically significant increase in gasoline price in the ARCO gas stations? Be very careful

(c) What is the key condition for the validity of this estimator? What data would you collect and what test would you do to support its validity in this case?