EEP 118 / IAS 118 Elisabeth Sadoulet and Ben Crost
University of California at Berkeley Fall 2009

Introductory Applied Econometrics
Final examination

There are 26 sub-questions, and they all have an equal weight of 5 points for a total of 130 points.

Your name:____ SID:__________________________

SID ________________________________

1. Below is the estimation of a standard model of household energy consumption:

$$\widehat{\log}(energy) = \hat{\beta}_0 + \hat{\beta}_1 \log(income) + \hat{\beta}_2 \log(price)$$

$$= \underset{(0.57)}{2.31} + \underset{(0.5)}{1.18} \log(income) - \underset{(0.25)}{0.60} \log(price) \qquad R^2 = .55 \; n = 250$$

where *energy* is the annual consumption of energy of the household (in 1000 kilowatt-hours), *income* its annual income (in \$1,000), and *price* the average price of energy over the year (in cents/kilowatt-hours), and standard errors are in parentheses.

a. What is the economic interpretation of the model coefficient β_2 ?

b. Interpret the estimated coefficient $\hat{\beta}_2$.

c. What is the predicted effect of a 15% increase in the price of energy on household average energy consumption?

d. Interpret the value of R^2

2. An equation for the revenue of a store is estimated from 235 weekly observations on *revenue* (in $1,000), average *price* of the products that it sells (in $), and its expenditures on *advertising* (in $1,000):

$$\underset{(6.6)}{revenue = 104.8} - \underset{(3.46)}{6.58\,price} + \underset{(0.79)}{1.95\,advertising} - \underset{(0.03)}{0.09\,advertising^2} \qquad R^2 = .87 \; n = 235$$

a. Is the marginal return to advertising increasing, decreasing, or constant as expenditures in advertising increase?

b. What is the optimal level of advertising for the store?

3. An equation for CEO (chief executive officer) salary estimated on 209 firms is:

$$\log(salary) = \underset{(0.30)}{4.59} + \underset{(0.032)}{.257\log(sales)} + \underset{(0.004)}{.011\log(roe)} + \underset{(0.089)}{.158\,finance} + \underset{(0.085)}{.181\,consprod} + \underset{(0.099)}{.283\,utility} \qquad R^2 = .357$$

where the variables are defined as follows: *salary* is 1990 salary in thousands of $, *sales* is 1990 firm sales in millions of $, *roe* is average 1988-90 return on equity, and *finance, consprod,* and *utility* are dummy variables indicating that the firm is in the finance, consumer product, or utility industry, respectively. The omitted industry category is transportation.

a. Use the above equation to test the hypothesis (at the 1% significance level) that return on equity has no effect on CEO salary

b. Use the above equation to test the hypothesis (at the 5% significance level) that the elasticity of CEO salary with respect to firm sales is equal to one.

c. Construct a 95% confidence interval for the elasticity of CEO salary with respect to firm sales.

d. Now you re-estimate the above equation with *finance*, *consprod*, and *utility* omitted. You obtain:

$$\log(salary) = \underset{(0.29)}{4.36} + \underset{(0.033)}{.275}\log(sales) + \underset{(0.004)}{.018}\log(roe) \qquad R^2 = .282$$

Use these two equations to test the joint hypothesis that CEO salary is unaffected by industry type.

4. You have a data set that contains information about individuals' gender, the number of children they have, and whether they are in the labor force

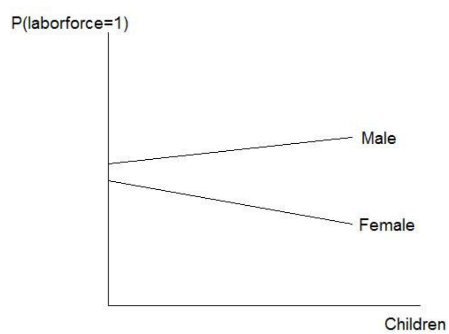
female	=1 if the individual is female, 0 if male
laborforce	=1 if the individual is in the labor force, 0 otherwise
children	number of children the individual has

You estimate the following linear probability model:

$$P(laborforce=1) = \beta_0 + \beta_1 children + \beta_2 female + \beta_3 (children \times female) + u$$

a. In terms of the model's parameters, what is the marginal effect of having an additional child on a woman's probability of being in the labor force? What is the marginal effect of having an additional child on a man's probability of being in the labor force?

b. Based on the graph below, what signs do you expect for the parameters β_0, β_1, β_2, and β_3?



5. Suppose the Berkeley Unified School District introduces a new after-school tutoring program, while the Oakland Unified School District doesn't. For both school districts, you observe the average score on a standardized Math exam in the year before and the year after Berkeley introduces the program:

Average score (on a maximum of 100)

	Berkeley	Oakland
Before program	60	55
After program	70	58

a. Calculate the difference-in-differences estimate of the effect of the tutoring program.

b. Suppose you estimate the following model for a student exam result:

$$score = \beta_0 + \beta_1 berkeley + \beta_2 after + \beta_3 berkeley{*}after + u$$

where *score* is the score obtained by the student on the math exam, *berkeley* is a dummy variable equal to one if the student attends a Berkeley school, 0 if he attends an Oakland school, and *after* is a dummy variable equal to one for the test taken after Berkeley introduces the program, and 0 for the test taken before. Which parameter yields the difference-in-difference estimate of the effect of the tutoring program?

c. What estimates do you expect for the parameters β_0, β_1, and β_2 in the regression in part b.?

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 ez_{it} is a dummy variable equal to 1 if the city *i* has an enterprise zone in year *t*, the variable $unclms_{it}$ is the number of unemployment claims filed during year *t* in city *i*, and *d*81, *d*82, …, *d*88 are dummy variables for the years 1981 to 1988.

a. Does the following model A allow you to identify a causal effect of having an enterprise zone on unemployment? Why or why not?

$$\log\left(uclms_{it}\right) = \beta_0 + \beta_1 ez_{it} + u_{it} \qquad \text{(Model A)}$$

b. Write the models whose estimations are reported in the following two Stata output. [Be very careful with indices].

c. What does model B control for that was a possible source of bias in estimating the causal effect of *ez* with model A?

d. What does model C control for that was a possible source of bias in estimating the causal effect of *ez* with model A?

e. Interpret the coefficients of *d*82 and *ez* in estimated model C.

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
```

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

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
```

Continuation of 7.

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 is enrollment affected by the gender of the child?

c. Using the results of model B, how does the distance to school affect the probability of school enrollment?

8. For the U.S. economy, let *lprice* denote the logarithm of the overall price level and *lwage* the logarithm of the average monthly wage of workers. Using data from January 1964 through October 1987, you estimated the following regression:

$$lprice_t = 0.002 + 0.19 lwage_t + 0.18 lwage_{t-1} + 0.11 lwage_{t-2}$$
$$t-stat: \quad (5.9) \quad (4.8) \qquad (4.6) \qquad (2.9)$$

a. Suppose that workers are given a bonus (i.e., a temporary wage increase) of 10 percent on their wage for just one month in December. What is the effect of this wage bonus on the predicted overall price level in December and in the following months?

b. Now suppose that the workers obtain a permanent increase of 10 percent in their wage. What is the effect of this increase on the predicted overall price level in December and in the following months?

c. What can you conclude on the long-term effect of this permanent increase in wage on the *real* wage of the workers? [Recall that the real wage is the wage divided by the price index]

Formulae

Statistics and miscellaneous

Covariance between two variables in a population: $\text{cov}(x,y) = \frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})$

$\text{cov}(a_1 x + b_1, a_2 y + b_2) = a_1 a_2 \text{ cov}(x,y)$

$\text{var}(x+y) = \text{var } x + \text{var } y + 2\text{ cov}(x,y)$

When y is a binary variable with probability prob(y = 1) = $p(x)$, the variance conditional on x is $p(x)(1-p(x))$

For small values of x: $e^{ax} \approx 1 + ax$

OLS estimator

$\hat{\beta}_1 = \frac{\text{cov}(x,y)}{\text{var } x}$ with $\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{SST_x}$

For multiple regression: $\text{var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}$

Adjusted R square: $\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\hat{\sigma}^2}{SST/(n-1)}$

Test statistics:

Loglikelihood ratio statistic for q restrictions: $LR = 2(Loglikelihood_{UR} - Loglikelihood_R) \sim \chi_q^2$

F statistic for q restrictions in a regression done with n observations and k exogenous variables:

$$\frac{(R_{UR}^2 - R_R^2)/q}{(1-R_{UR}^2)/(n-k-1)} \sim F(n-k-1, q)$$

Chow statistic: $F = \frac{\left[SSR_P - (SSR_1 + SSR_2)\right]/k+1}{SSR_1 + SSR_2/\left[n-2(k+1)\right]} : F(k+1, n-2(k+1))$