EEP 118 / IAS 118 Elisabeth Sadoulet and Daley Kutzman
University of California at Berkeley Fall 2013

Introductory Applied Econometrics
Midterm examination

Scores add up to 50 (5 points for each sub-question)

Your name:_____ SID:_____

1. (5 points) Using data on birth weights, we estimated the following two models:

$$\widehat{bwght} = 119.8 - 0.52\,cigs \qquad R^2=0.023 \quad n=1388$$
$$(0.6) \quad (0.09)$$

$$\widehat{bwght} = 116.9 - 0.46\,cigs + 0.09\,faminc \qquad R^2=0.03 \quad n=1388$$
$$(1.0) \quad (0.09) \quad (0.03)$$

where *bwght* is birth weight (in ounces), *cigs* is the number of cigarettes smoked daily during pregnancy and *faminc* is 1988 family income, in $1000.
How does the introduction of the variable *faminc* affect the estimated parameter on *cigs*? What can you infer about the correlation between *faminc* and *cigs*? Justify your response.

2. (5 points) Below are summary statistics for the GPAs of a sample of 101 students from Michigan State University. The dean of MSU, Lou Anna Simon, firmly believes that the true average GPA of her university is 3.1 and your sample below is an inaccurate representation. Should you be skeptical of Lou Anna's claim? Support your opinion with a hypothesis test at the 5% significance level.

Variable	Obs	Mean	Std. Dev.	Min	Max
colGPA	101	2.984	.3723103	2.2	4

3. (10 points) Using a small sample of households from Nicaragua, we estimate the relationship between the log of energy expenditure (`lenergyexp`) and the log of household total expenditure per capita (`ltotexppc`), household size (`hhsize`), and whether the household owns a stove (`stove`).

```
. regress lenergyexp ltotexppc lhhsize stove
```

Source	SS	df	MS		
Model	82.5864529	3	27.5288176	Number of obs	= 174
Residual	155.493187	170	.914665804	F(3, 170)	= 30.10
				Prob > F	= 0.0000
				R-squared	=
Total	238.07964	173	1.37618289	Adj R-squared	= 0.3354
				Root MSE	= .95638

lenergyexp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ltotexppc	.7763713	.1014082	7.66	0.000	.5761897	.9765528
lhhsize	.4904821	.1654038	2.97	0.003	.1639723	.816992
stove	.358545	.2240303				
_cons	-.7896446	1.091029	-0.72	0.470	-2.943354	1.364064

a. What is the p-value for the test that having a stove has no effect on the consumption of energy? Interpret your results.

b. Calculate and interpret the R-squared for this estimation.

4. (15 points) You have n = 40 quarterly observations on the imports M of a country, an index of import prices P_M, and real aggregate income GDP. Adding dummy variables $Q2, Q3$, and $Q4$ for the 2nd, 3rd, and 4th quarters of the year, you estimate the model:

$$\log M = \beta_0 + \beta_1 \log P_M + \beta_2 \log GDP + \beta_3 Q2 + \beta_4 Q3 + \beta_5 Q4 + u$$

and find the following results:

$$\widehat{\log M} = 4.30 - 0.58 \log P_M + 1.45 \log GDP + .15 Q2 + .10 Q3 + .40 Q4 \qquad R^2 = 0.253,\ n = 40$$

(0.13) (0.21) (.10) (.05) (.12)

(a) Construct a 95% confidence interval for β_1. Interpret.

(b) Test the hypothesis $\beta_2 = 1$ against $\beta_2 \neq 1$ at the 5% significance level. Interpret this result in economic terms.

(c) Why is a first quarter $Q1$ dummy variable not included in the model? Interpret the estimated parameters on $Q2, Q3$, and $Q4$.

5. (15 points) Using data for the US gasoline market between 1960 and 1999, we estimated the following model:

```
. regress lng lninc lnpriceg lnprnewcar lnprusedcar
      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  4,    35) =  440.79
       Model |  2.14671037     4   .30667291           Prob > F      =  0.0000
    Residual |  .019480675    35  .000695738           R-squared     =  0.9910
-------------+------------------------------           Adj R-squared =  0.9888
       Total |  2.16619104    39  .061891173           Root MSE      =  .02638

------------------------------------------------------------------------------
         lng |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lninc |   1.692181   .2154954     7.85   0.000     1.250759    2.133604
    lnpriceg |  -.2325466   .0595617    -3.90   0.001    -.3545532   -.1105399
  lnprnewcar |   -.233414   .1795357    -1.30   0.204    -.6011761    .1343482
 lnprusedcar |  -.0597329   .0693655    -0.86   0.396    -.2018216    .0823558
       _cons |  -3.957868   2.581538    -1.53   0.136     -9.24591    1.330174
------------------------------------------------------------------------------
```

where: `lng`=log(total US gasoline consumption per capita), `lninc` = log(income per capita), `lnpriceg` = log(gasoline price), `lnprnewcar` = log(price of new cars), and `lnprusedcar` = log(price of old cars)

a. Suppose the government imposes a tax on gasoline that induces a price increase of 15%. What would be the effect on gasoline consumption?

b. Is the consumption of gasoline influenced by the price of new cars and the price of used cars, when you consider them one at a time? Justify your response.

c. We now estimate the model without the prices of new and used cars (See table on the last page). Comparing the two estimated models, would you say that the consumption of gasoline is influenced by the prices of cars, new or used, considered together? (Do a joint test of significance on the two parameters).

Formulae

Statistics

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

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

$\operatorname{var}(ax + by) = a^2 \operatorname{var} x + b^2 \operatorname{var} y + 2ab \operatorname{cov}(x,y)$

Variance for the difference in means of two independent samples:
$\operatorname{var}(\bar{x}_1 - \bar{x}_2) = \operatorname{var}(\bar{x}_1) + \operatorname{var}(\bar{x}_2)$

When y is a binary variable with probability prob(y = 1) = p, its variance is p (1– p)

OLS estimator

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

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

$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$, $SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$, and $SSR = \sum_{i=1}^{n} \hat{u}_i^2$

Test statistics:

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

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

Table for question 5c

```
. regress lng lninc lnpriceg
      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  2,    37) =  629.99
       Model |  2.13986669     2  .534966673           Prob > F      =  0.0000
    Residual |  .026324354    37  .000849173           R-squared     =  0.9878
-------------+------------------------------           Adj R-squared =  0.9863
       Total |  2.16619104    39  .061891173           Root MSE      =  .02914

------------------------------------------------------------------------------
         lng |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lninc |    2.13019    .148784    14.32   0.000     1.826743    2.433637
    lnpriceg |  -.1528558   .0535811    -2.85   0.008    -.2621352   -.0435763
       _cons |  -5.605895   2.165544    -2.59   0.015    -10.02255    -1.18924
------------------------------------------------------------------------------
```