Econometrics Preliminary Examination

Department of Agricultural and Resource Economics
University of California, Davis

July 3, 2013

Answer each part of each question. All questions are weighted equally. Within each question, each part will receive equal weight in grading.

Question One

a. A gambler has two coins in her pocket. One coin shows Heads on both sides; the other is a fair coin. She picks one coin at random and looks at one side of the coin. What is the probability that she sees Heads? Given that she sees Heads, what is the probability that the other side shows Heads?

b. How do your answers in (a) change if she has one two-headed coin and four fair coins in her pocket?

c. The gambler would like the conditional probability that the other side shows Heads, given that the first side does so, to equal 1/10. How many fair coins should she have in her pocket?

Question Two

Suppose you observe n independent and identically distributed Gamma random variables, $Y_1, Y_2, \ldots, Y_n$. Recall that the Gamma has a probability density function given by

$$f(y) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} e^{-y/\beta} y^{\alpha-1} & y > 0 \\ 0 & otherwise \end{cases}$$

Both α and β are unknown.

Your research assistant misunderstood your instructions concerning the data, and deleted the spreadsheet containing the n observations. You have only two pieces of information, from Stata's summarize command: the sample mean of the n observations for Y and the sample mean of the observed values of Y^2. Both of these sample means were calculated using all n observations.

a. You would like to estimate the two parameters α and β using the information you have available, so you decide to consider whether asymptotic theory provides any guide.

What are the probability limits of your two sample statistics?

b. Based on your answer to part *(a)*, explain how you can use the two sample statistics to estimate the two parameters of interest. What properties are you able to establish for your two estimators?

Question Three

Let X denote the lifetime of a light bulb manufactured using one manufacturing process, by a certain company, and let Y denote the lifetime of a light bulb manufactured using an alternative manufacturing process, by the same company. You believe that X is distributed as an Exponential random variable with mean β_1 and that Y is distributed as an Exponential random variable with mean β_2.

Recall that the Exponential random variable has a probability density function given by

$$f(y) = \begin{cases} \frac{1}{\beta} e^{-y/\beta} & y > 0 \\ 0 & otherwise \end{cases}$$

Explain how you could use the Central Limit Theorem, along with a large number of observations for each random variable, to test the hypothesis that X and Y are identically distributed Exponential random variables.

Question Four

Given the generalized linear regression model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$
$$E[\mathbf{u} \mid \mathbf{X}] = \mathbf{0}$$
$$E[\mathbf{u}\mathbf{u}' \mid \mathbf{X}] = \sigma^2\Omega = \Sigma$$

a. Show that the OLS estimator of β is unbiased.

b. Find the variance-covariance matrix for the OLS estimator of β, either assuming that the regressors are non-stochastic or conditional on $\mathbf{X}$.

c. Show that the OLS estimator of β is consistent.

d. Are the properties of the OLS estimator of β the same if the disturbance term exhibits first-order autocorrelation? Explain.

e. Indicate whether your answer to (d) changes, for the case where the autocorrelation is a result of omitted variables.

Question Five

For each of the following questions, answer True or False and give a justification for your response.

a. For the regression equation in question 4, $E(u_i) = 0$ implies that

$$\sum_{i=1}^{n} u_i = 0$$

where n denotes the sample size.

b. For the regression equation in question 4, $E(u_i) = 0$ implies that

$$\sum_{i=1}^{n} e_i = 0$$

where e_i denotes the fitted residual for observation i.

c. The sum of the least squares residuals equals zero, i.e.

$$\sum_{i=1}^{n} e_i = 0$$

and this implies that $R^2 = 1$.

d. $E(uu') = \sigma^2 I$ implies that the least squares residuals are uncorrelated with each other.

e. The finite-sample properties of an estimator are the same as the asymptotic sampling properties.

f. Consistency implies that that an estimator is also asymptotically efficient.

g. An estimator that is asymptotically unbiased is also consistent.

h. If a random variable converges in probability to a, then it must be true that the limit of the expected value of the random variable equals a.

Question Six

Suppose you specify the econometric model $Y_i = \alpha + \beta X_i + \varepsilon_i$, where ε_i is normally distributed with mean zero. Consider the following set of results generated from a random sample of individuals:

Variable	Mean	Std.Dev.
X	49.51	28.74
Y	51.22	59.25

OLS Estimates	
X	1.02**
	(0.06)
Constant	0.61
	(3.74)
R^2	0.25
N	1000

Standard errors in parentheses
** Significant at 95% confidence level
N is the number of observations

In answering the following questions, justify your answers and state any additional assumptions you make.

a. What would be the minimum value of X that an individual would need to make sure that s/he obtains a positive outcome (Y > 0) with 95% probability?

b. What minimum, positive value of X would make sure, with 95% probability, that the individual obtains more Y than a person who has X = 0?

c. Given that the 95% confidence interval for β is (0.90, 1.14), if an individual has X = 1, what would be the probability that s/he gets Y > 0.90?

Question Seven

In a 1996 QJE paper, Steven Levitt estimates the effect of imprisonment on violent crime. Consider two linear models:

$$gcriv = \alpha_0 + \alpha_1 gpris + \alpha_2 gincpc + u, \qquad E[u \mid gincpc, gpris] = 0$$

$$gcriv = \beta_0 + \beta_1 gpris + \beta_2 gincpc + \varepsilon, \qquad E[\varepsilon \mid gincpc, final1, final2] = 0$$

where gcriv denotes the annual growth rate in violent crime

gpris denotes the annual growth rate in the number of prison inmates per resident

gincpc denotes per capita income

final1 is a dummy variable denoting a final decision on prison overcrowding legislation in the current year

final2 is a dummy variable denoting a final decision on prison overcrowding legislation in the previous two years.

The data are measured annually at the state level, i.e., one observation for each U.S. state in each year from 1980-1993.

a. Suppose that (i) high violent crime rates cause more people to be imprisoned, (ii) decisions on prison overcrowding legislation are random, and (iii) decisions on prison overcrowding legislation cause reductions in the prison population. Would you expect α_1 to exceed β_1 or vice versa? Explain in words.

b. Suppose prison overcrowding legislation is less likely to reach a final decision in high-violent-crime years because politicians do not want to be perceived as allowing more criminals to be free. What are the implications of this possibility for your interpretation of β_1?

Your research assistant brings to you output from three regressions (see the last three pages of the exam).

c. Following regression II, your research assistant used the “estat firststage” command to generate a first stage *F* statistic of 24.69. Explain how this statistic was computed and what you conclude from it about the properties of regression II.

d. Following regression II, your research assistant used the “estat endogenous” command to conduct a Hausman test. Explain how this statistic was computed and what you conclude from it about the properties of regression II.

e. Interpret the coefficient on *gpris* in regression III and contrast it with the coefficient on *gpris* in regression I.

f. Would you make any recommendation to your research assistant to modify the method used to estimate the standard errors? Why or why not?

REGRESSION OUTPUT FOR QUESTION SEVEN

REGRESSION I

```
. reg gcriv gpris gincpc, robust

Linear regression                                      Number of obs =     714
                                                       F(  2,   711) =   15.62
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0461
                                                       Root MSE      =  .08661

------------------------------------------------------------------------------
             |               Robust
       gcriv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gpris |  -.1954158   .0531582    -3.68   0.000    -.2997817    -.09105
      gincpc |   .4667109   .1468838     3.18   0.002     .1783331    .7550887
       _cons |   .0043254   .0106637     0.41   0.685    -.0166106    .0252615
------------------------------------------------------------------------------
```

REGRESSION II

```
. ivregress 2sls gcriv (gpris=final1 final2) gincpc, robust first

First-stage regressions
-----------------------
                                                  Number of obs   =        714
                                                  F(   3,    710) =      16.48
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.0273
                                                  Adj R-squared   =     0.0232
                                                  Root MSE        =     0.0658

------------------------------------------------------------------------------
             |               Robust
       gpris |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gincpc |  -.1032158   .1211331    -0.85   0.394    -.3410377    .1346061
      final1 |  -.0743413   .0117319    -6.34   0.000    -.0973747   -.0513078
      final2 |  -.0643859   .0196345    -3.28   0.001    -.1029345   -.0258373
       _cons |   .0743395   .0071673    10.37   0.000     .0602678    .0884112
------------------------------------------------------------------------------
Instrumental variables (2SLS) regression               Number of obs =     714
                                                       Wald chi2(2)  =   17.72
                                                       Prob > chi2   =  0.0001
                                                       R-squared     =       .
                                                       Root MSE      =  .10462

------------------------------------------------------------------------------
             |               Robust
       gcriv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gpris |  -1.082207   .3147788    -3.44   0.001    -1.699163   -.4652521
      gincpc |   .3798519   .2003237     1.90   0.058    -.0127754    .7724792
       _cons |   .0684567   .0255346     2.68   0.007     .0184098    .1185036
------------------------------------------------------------------------------
Instrumented:  gpris
Instruments:   gincpc final1 final2
```

```
. estat firststage

  First-stage regression summary statistics
  --------------------------------------------------------------------------
               |            Adjusted      Partial       Robust
      Variable |   R-sq.       R-sq.        R-sq.      F(2,710)   Prob > F
  -------------+------------------------------------------------------------
         gpris |  0.0273      0.0232       0.0256       24.6875    0.0000
  --------------------------------------------------------------------------


. estat endogenous

  Tests of endogeneity
  Ho: variables are exogenous

  Robust score chi2(1)            =  6.86002  (p = 0.0088)
  Robust regression F(1,710)      =  10.5555  (p = 0.0012)

. estat overid

  Test of overidentifying restrictions:

  Score chi2(1)          =  .032366  (p = 0.8572)
```

REGRESSION III

```
. reg gcriv gpris gincpc final1 final2, robust

Linear regression                                      Number of obs =     714
                                                       F(  4,   709) =   12.52
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0578
                                                       Root MSE      =   .0862

------------------------------------------------------------------------------
             |               Robust
       gcriv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gpris |  -.1721091   .0531589    -3.24   0.001    -.2764768   -.0677414
      gincpc |   .4752465   .1449376     3.28   0.001     .1906883    .7598047
      final1 |    .072938   .0430142     1.70   0.090    -.0115124    .1573884
      final2 |    .055544   .0161444     3.44   0.001     .0238475    .0872405
       _cons |   .0007204   .0105788     0.07   0.946    -.0200492      .02149
------------------------------------------------------------------------------

.
```

Econometrics Preliminary Examination

Department of Agricultural and Resource Economics
University of California, Davis

August 19, 2013

Answer each part of each question. All questions are weighted equally. Within each question, each part will receive equal weight in grading.

If you feel the need for a calculator, note that in such circumstances, exact answers are not required. Please state any approximating assumptions you make.

Question One

A coin with Heads on one side and Tails on the other is flipped an unknown number of times (n), and the number of Heads is observed. *You do not know whether or not the coin is fair.* This experiment will be repeated N times, where N is known to you. The only information you will have is the observed outcomes for the number of Heads each time the experiment is performed: $Y_1, Y_2, \ldots, Y_N$. Once the N observations are available, you will calculate the sample mean of your Y_i observations.

Suppose you plan to repeat the entire procedure M times. Describe the probability distribution (including expressions for the expected value and variance) that you expect your M observed sample means to follow. Distinguish carefully between small and large values for M, and answer for both cases.

Question Two

A random variable Y is known to follow the Normal distribution:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-(y-\mu)^2/\sigma^2} \quad -\infty < y < +\infty$$

You will have n observations from this distribution. Both parameter values are unknown. Find maximum-likelihood estimates for μ and σ^2. Explain why your estimate of σ^2 is biased, and give an unbiased estimator. Prove that your new estimator is unbiased, and prove that your new estimator and your biased, maximum-likelihood estimator for σ^2 are both consistent.

Question Three

Which of the following can cause OLS estimators to be biased? Give a brief justification for each answer.

(i) Heteroskedasticity
(ii) Omitting an important variable
(iii) A sample correlation coefficient of 0.96 between two explanatory variables both included in the model
(iv) Autocorrelation
(v) Autocorrelation when the lagged dependent variable appears as an explanatory variable

Question Four

a. A person states that if a researcher is testing the equality between two means and the sample size is sufficiently large, then a statistically significant result can always be obtained. True or false? Explain.
b. Since every model is misspecified by not including one or more relevant variables, then ordinary least squares estimators are always biased and inconsistent. Comment.
c. Give five examples of things that can affect the power of statistical tests.
d. Give an example of sample selectivity bias and how it might impact regression estimates.

Question Five

Spending on medical care varies widely across counties in the United States, as does life expectancy. Suppose you are interested in estimating the causal effect of medical-care spending on life expectancy.

You begin by specifying the model:

$$LE_i = \beta_0 + \beta_1 M_i + \varepsilon_i$$

where LE_i denotes life expectancy in county i and M_i denotes spending per person on medical care in county i.

a. Suppose some counties in your sample have healthier populations than other counties, which causes them to have higher life expectancies. Given this information, what will be the properties of the OLS estimate of β_1 in the above regression model?

b. Continuing from (a), suppose you observe for each county a variable X that measures the health of the population. How would you incorporate X into your regression model? Justify your answer.

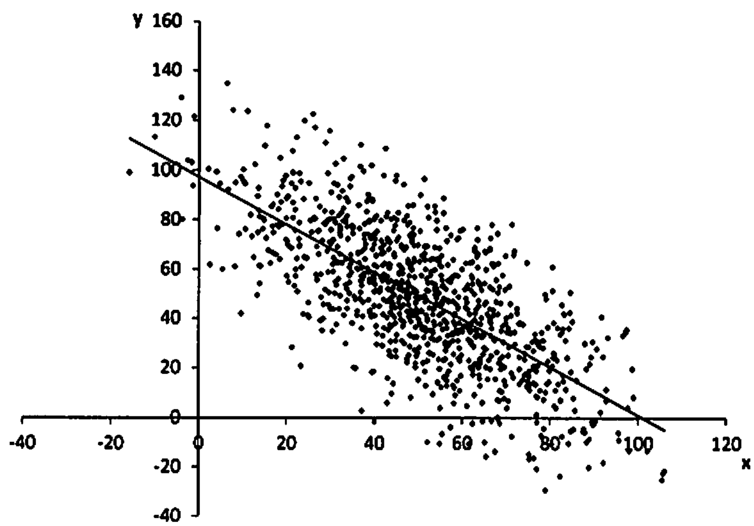
c. Continuing from (a), suppose you observe an instrumental variable Z. What properties must the instrumental variable possess for it to be useful in producing a consistent estimate of β_1 with a small bias?

d. Suppose you learn that your data do not come from the United States as we know it. Instead, they come from an imaginary country that assigns medical spending to counties randomly (i.e., half of counties are assigned high spending levels and the rest are assigned low spending). Given this information, how would you estimate β_1? Justify your answer. You may assume that you observe the variables Z and X from parts (b) and (c).

e. Continuing from (d), suppose that in spite of the government assignment of spending levels, many counties spend a different amount on medical care than they are assigned. Explain how you would test whether these deviations cause bias in your estimate of β_1. Write down the test statistic you would use and state its asymptotic null distribution. You may assume that you observe the variables Z and X from parts (b) and (c).

Question Six

Suppose you specify the econometric model $Y_i = \alpha + \beta X_i + \varepsilon_i$, where ε_i is normally distributed with mean zero and a constant variance. You have data on a random sample of 1000 individuals. Here is a scatter plot of the data, including the least squares regression line.



Here are the results from an OLS regression

Variable	Mean	Std.Dev.
X	50.41	20.05
Y	48.71	27.46

	OLS Estimates
X	-0.96**
	(0.03)
Constant	97.25**

	(1.67)
R^2	0.50
N	1000

Standard errors in parentheses
** Significant at 95% confidence level
N is the number of observations

In answering the following questions, justify your answers and state any additional assumptions you make.

a. What is the maximum value of X such that an individual has a 95% probability of a positive outcome (Y > 0)?

b. What is the maximum value of X such that, with 95% probability, the individual obtains more Y than a person who has X = 100?

c. Consider an individual with X=99 and another individual with X=100. What is the expected difference in Y between these two individuals?

d. Give a 95% confidence interval for *expected difference* you computed in (c).