Agricultural and Resource Economics

Ph.D. Qualifying Examination – Quantitative Methods

Monday, July 10, 2017

Instructions:

1) You will have 3.5 hours to complete the exam.
2) There are five questions on five pages. You must answer all five questions.
3) You may not use outside resources, including textbooks, notes, calculators, or electronic devices.
4) Show all your work on the paper provided.
5) Leave one-inch margins and only write on one side of each sheet.
6) Do not place your name on any of your answer pages.
7) Clearly number each sheet with the question and page number in the upper right corner.
8) Submit your exam unstapled ordered by question number and page number.
9) If you need clarification about a question or believe there is a typographical error, raise your hand and the exam invigilator will assist you.
10) You may consume drinks and/or snacks, as long as doing so does not distract other students.
11) Students may use the restroom if they inform the invigilator, but only one student may be absent from the examination room at a time.
12) Students may not leave the examination early.

1. A researcher estimates the following regression equation with one thousand observations (the data have been demeaned to remove the constant), where q_i denotes quantity demanded of good *i*, p_i denotes the price of good *i* and *m* denotes total expenditure:

$$\log q_1 = \alpha \log m + \beta \log p_1 + \gamma \log p_2 + \varepsilon_i$$

They report the following coefficient estimates and covariance matrix.

$$\begin{aligned} \alpha &= 0.59 \\ \beta &= -0.92 \\ \gamma &= 0.70 \end{aligned}$$

$$\Omega = \begin{bmatrix} .0009 & .0015 & .0002 \\ & .0036 & -.0050 \\ & & .0100 \end{bmatrix}$$

A. Can we reject the null hypothesis that demand for good 1 is inelastic at the five percent significance level?

B. Can we reject the null hypothesis that demand for good 1 satisfies homogeneity at the five percent significance level?

2. The true relationship that determines outcome y is:

$$y_i = \beta x_i + \gamma x_i z_i + \varepsilon_i$$

where y_i, x_i, z_i and ε_i are scalars and the data have been demeaned to remove the constant. An econometrician lacks the variable z in their dataset and thus estimates the following equation using OLS:

$$y_i = \delta x_i + \varphi_i$$

A. Calculate the probability limit of $\hat{\delta}$, the OLS estimate of δ, assuming that $x \perp \varepsilon$

B. Suppose that x is distributed uniform on the interval [-1,1] and z is distributed uniform on the interval [-2,2]. Compare the probability limit of $\hat{\delta}$ to β.

3. A researcher wishes to recover the causal effect of child mortality on subsequent fertility. The researcher has access to a dataset that records the following information for households in a large tropical nation:

Maternal age
Number of children
Household income
Maternal years of education
Township of residence
Month and year of interview
Indicator for death of a child in the past year
Indicator for birth of a child in the past year
Indicator for current pregnancy status

3.a. Explain why estimating the relationship between current pregnancy status and death of a child in the past year using OLS (a linear probability model) might yield inconsistent estimates of the causal effect of mental health on employment.

The researcher proposes the following instrumental variable for the death of a child in the past year: an index of malaria transmission danger. This index is constructed based on ecological and climatic factors using a method validated in the epidemiological literature.

3.b. List each of the assumptions necessary for the IV estimate to be interpreted as the LATE of child mortality on subsequent fertility when malaria transmission risk is used as an instrument. For each assumption, explain intuitively what is required for the proposed instrument to meet the stated requirement.

3.c. For each of the assumptions, provide an intuitive example of when the proposed instrument may fail that requirement.

4. For a sample of 1755 white males in the United States in the year 2000, a researcher has the following data:

> *years of schooling*, ***S***,
> *weekly earnings in dollars*, ***W***,
> *hours worked per week*, ***H***, and
> *hourly earnings*, ***E*** (computed as *W*/*H*)

She calculates *LW*, *LE*, and *LH* as the natural logarithms of *W*, *E*, and *H*, respectively, and reports the following results:

Table 2: OLS Regressions Results

	(1)	(2)	(3)	(4)	(5)
Respondents	All	All	All	Full-time	Part-time
Dependent Variable	LE	LW	LE	LW	LW
S	0.099 (0.006)	0.098 (0.006)	0.098 (0.006)	0.101 (0.006)	0.030 (0.049)
LH	-	1.190 (0.065)	1.190 (0.065)	0.980 (0.088)	0.885 (0.325)
Constant	6.111 (0.082)	5.403 (0.254)	5.403 (0.254)	6.177 (0.345)	7.002 (1.093)
RSS	741.5	737.9	737.9	626.1	100.1
Observations	1755	1755	1755	1669	86

Notes: Standard errors in parentheses; RSS = residual sum of squares. Full-time employment defined as working at least 35 hours per week. The correlation between ***S*** and ***LH*** is 0.06.

(a) Explain why specification (1) is a restricted version of specification (2), stating and interpreting the restriction.

(b) Supposing the restriction to be valid, explain whether you expect the coefficient of *S* and its standard error to differ, or be similar, in specifications (1) and (2).

(c) Supposing the restriction to be invalid, how would you expect the coefficient of *S* and its standard error to differ, or be similar, in specifications (1) and (2)?

(d) Perform an *F* test of the restriction.

(e) Perform a *t* test of the restriction.

(f) Explain whether the *F* test and the *t* test could lead to different conclusions.

(g) Interpret the results reported in columns 2 and 3. Why are some estimates the same and others different?

(h) At a seminar, a commentator says that part-time workers tend to be paid worse than full-time workers and that their earnings functions are different. Based on the results reported above, test whether the commentator's assertion is correct. (Hint: For a reference, the critical value of $F(3,1000)$ at the 0.1% level is 5.46.)

(i) What are the implications of the commentator's assertion for the test of the restriction?

5. Answer the following questions:

a.) Suppose that a variable Y depends on a variable X with the relationship

$$Y = \beta_2 X + u$$

where u is a disturbance term that satisfies the Gauss–Markov conditions.

(i) Explain in principle how one would derive the ordinary least squares (OLS) estimator of β_2.

(ii) Demonstrate that for this model the OLS estimator is $\frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$

(iii) Explain in general terms why OLS is an attractive estimation procedure if the disturbance term in the model satisfies the Gauss–Markov conditions. *Note*: Mathematical proofs of any assertions are not required.

b.) A variable Y depends on a nonstochastic variable X with the relationship

$$Y = \beta_1 + \beta_2 X + u$$

where u is a disturbance term that satisfies the Gauss–Markov conditions. You may assume that $\frac{Cov(X,Y)}{Var(X)}$ is the OLS estimator of β_2 for this model, that it is unbiased, and that it has population variance of $\frac{\sigma_u^2}{nVar(X)}$.

Given a sample of n observations, a researcher decides to estimate β_2 using the expression

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

derived in part (a). It can be shown that the population variance of this estimator is $\frac{\sigma_u^2}{\sum_{i=1}^{n} X_i^2}$.

(i) Demonstrate that $\hat{\beta}_2$ is in general a biased estimator of β_2.

(ii) Discuss whether it is possible to determine the sign of the bias.

(iii) Demonstrate that $\hat{\beta}_2$ is unbiased if $\beta_1 = 0$. What can be said in this case about the efficiency of $\hat{\beta}_2$, comparing it with the alternative estimator $\frac{Cov(X,Y)}{Var(X)}$?

(iv) Demonstrate that $\hat{\beta}_2$ is unbiased if $\bar{X} = 0$. What can be said in this case about the efficiency of $\hat{\beta}_2$, comparing it with the alternative estimator $\frac{Cov(X,Y)}{Var(X)}$? Explain the underlying reason for your conclusion.

(v) Returning to the general case where $\beta_1 \neq 0$ and $\bar{X} \neq 0$ and, suppose that there is very little variation in X in the sample. Is it possible that $\hat{\beta}_2$ might be a better estimator than the OLS estimator?

Agricultural and Resource Economics

Ph.D. Qualifying Examination – Quantitative Methods

Monday, June 13, 2016

Instructions:

1) You will have 3.5 hours to complete the exam.
2) There are four questions on four pages. You must answer all four questions.
3) You may not use outside resources, including textbooks, notes, calculators, or electronic devices.
4) Show all your work on the paper provided.
5) Leave one-inch margins and only write on one side of each sheet.
6) Do not place your name on any of your answer pages.
7) Clearly number each sheet with the question and page number in the upper right corner.
8) Submit your exam unstapled ordered by question number and page number.
9) If you need clarification about a question or believe there is a typographical error, raise your hand and the exam invigilator will assist you.
10) You may consume drinks and/or snacks, as long as doing so does not distract other students.
11) Students may use the restroom if they inform the invigilator, but only one student may be absent from the examination room at a time.
12) Students may not leave the examination early.

1) A researcher wishes to recover the causal effect of mental health on employment. The researcher has access to a dataset that records the following information for adults in the United States between 18 and 65 years of age:

number of poor mental health days in the past 30 days
number of days working at least 4 hours in the past 30 days
age
gender
race/ethnicity
household income
years of education
state and county of residence
county unemployment rate
month of interview

1.a. Explain why estimating the relationship between the number of poor mental health days and the number of days worked using OLS might yield inconsistent estimates of the causal effect of mental health on employment.

The researcher proposes the following instrumental variable for the number of poor mental health days: the average number of hours of daylight in the month of interview. Seasonal Affective Disorder (SAD) was a recognized mood disorder in the *Diagnostic and Statistical Manual of Mental Disorders-IV* associated with major depressive episodes. It is believed that a risk-factor for SAD is lack of sufficient natural daylight.

The amount of natural daylight is determined entirely by latitude and the tilt of earth's axis relative to its plane of revolution around the sun. It can thus be calculated for every respondent to the survey using the month of interview and county of residence.

1.b. List each of the assumptions necessary for the IV estimate to be interpreted as the LATE of poor mental health on employment when hours of daylight is the instrument. For each assumption, explain intuitively what is required for the proposed instrument to meet the stated requirement.

1.c. For each of the assumptions, provide an intuitive example of when the proposed instrument may fail that requirement.

2) You wish to estimate the following relationship using OLS:

$$y_i = x_i\beta + \varepsilon_i$$

where x_i is a scalar is a scalar for each individual, i=1,...,n. Assume the data have been demeaned to remove the constant and E[$x\varepsilon$]=0 .

Your data set does not include x_i. Instead, your data set includes $x_i^* = x_i + \eta_i$, where η_i is a mean-zero random variable and cov(ε_i, η_i)=0. Thus, you use OLS to estimate: $y_i = x_i^*\delta + \varepsilon_i$.

a) Derive the probability limit of $\hat{\delta}$, the OLS estimate of δ.

b) Suppose that E[$x\eta$]=0. Compare the probability limit of $\hat{\delta}$, to the true value of β.

c) Suppose that E[$x\eta$]>0 and β>0. Compare the probability limit of $\hat{\delta}$, to the true value of β.

d) What are the least restrictive conditions required for $\hat{\delta}$ to be a consistent estimator of β?

3) A researcher obtains data on household annual expenditure on books, B, and annual household income, Y, for 100 US households in 2006. She hypothesizes that B is related to Y and the average cognitive ability of adults in the household, IQ, by the relationship

$$\log B = \beta_1 + \beta_2 \log Y + \beta_3 \log IQ + u \qquad \text{(A)}$$

where u is a disturbance term that satisfies the Gauss–Markov conditions. She also considers the possibility that $\log B$ may be determined by $\log Y$ alone:

$$\log B = \beta_1 + \beta_2 \log Y + u \qquad \text{(B)}$$

She does not have data on IQ and decides to use average years of schooling of the adults in the household, S, as a proxy in specification (A). It may be assumed that Y and S are both nonstochastic. In the sample the correlation between $\log Y$ and $\log S$ is 0.86. She performs the following regressions:

(1) $\log B$ on both $\log Y$ and $\log S$, and
(2) $\log B$ on $\log Y$ only, with the results shown in the table (standard errors in parentheses):

	(1)	(2)
$\log Y$	1.10 (0.69)	2.10 (0.35)
$\log S$	0.59 (0.35)	–
constant	–6.89 (2.28)	–3.37 (0.89)
R^2	0.29	0.27

a. Assuming that (A) is the correct specification, explain, with a mathematical proof, whether you would expect the coefficient of $\log Y$ to be greater in regression (2).
b. Assuming that (A) is the correct specification, describe the various benefits from using $\log S$ as a proxy for $\log IQ$, as in regression (1), if $\log S$ is a good proxy.
c. Explain whether the low value of R^2 in regression (1) implies that $\log S$ is not a good proxy.
d. Assuming that (A) is the correct specification, provide an explanation of why the coefficients of $\log Y$ and $\log S$ in regression (1) are not significantly different from zero, using two-tailed t tests.
e. Discuss whether the researcher would be justified in using one-tailed t tests in regression (1).
f. Assuming that (B) is the correct specification, explain whether you would expect the coefficient of $\log Y$ to be lower in regression (1).
g. Assuming that (B) is the correct specification, explain whether the standard errors in regression (1) are valid estimates.

4) (a) A researcher believes that a model consists of the following relationships:

$$Y_t = \alpha_1 + \alpha_2 X_t + u_t \quad (1)$$

$$X_t = \beta_1 + \beta_2 Y_t + v_t \quad (2)$$

$$Z_t = \gamma_1 + \gamma_2 Y_t + \gamma_3 X_t + \gamma_4 Q_t + w_t \quad (3)$$

where u_t, v_t, and w_t, are disturbance terms that are drawn from fixed distributions with zero mean. It may be assumed that they are distributed independently of Q_t and of each other and that they are not subject to autocorrelation. All the parameters may be assumed to be positive and it may be assumed that $\alpha_2\beta_2 < 1$. When answering (i) – (iv) it should be assumed that the model is correctly specified.

i. Explain what is meant by an endogenous variable and an exogenous variable in a simultaneous equations model, and state which variables in this model are endogenous and exogenous.

ii. The researcher fits (2) using ordinary least squares (OLS). Evaluate whether the estimate of β_2 is likely to be biased. If it is biased, determine the direction of the bias.

iii. The researcher fits (3) using OLS. Determine whether the parameter estimates are likely to be biased. (You are not expected to evaluate the direction of the bias, if any.)

iv. The researcher decides to fit (2) using instrumental variables (IV), with Q_t as an instrument for Y_t. Determine whether he is likely to obtain a consistent estimate of β_2.

(b) At a seminar, it is agreed that a better specification for equation (2) would be

$$X_t = \beta_1 + \beta_2 Y_{t-1} + v_t \quad (2)^*$$

i. A participant at the seminar says that consistent estimates will be obtained if (2)* is fitted using OLS and (1) is fitted using IV, with Y_{t-1} as an instrument for X_t. Evaluate whether this is true.

ii. Another participant asserts that consistent estimates will be obtained if both (1) and (2)* are fitted using OLS, and that the estimate of α_2 will be more efficient than that obtained using IV. Evaluate whether this is true.

ARE – PHD Exam in Quantitative Methods, June 25, 2015

This exam is intended for 4 hours (9am – 1pm).

You may use snacks or drinks as you need to during the exam.

A professor is available in case there is a question of clarification, but he or she may decide to use the option of answering: "Do the best you can with the information provided."

1. ***Read the whole exam first***. Do not sign your name to the exam.

2. You must address ***at least*** 4 of the 5 questions completely, including all of the parts under a particular question. Addressing a 5th question may help make up for deficiencies (if any) in your answers to the 4 questions you emphasize.

3. Do your own work, without input from any other human by any means (including without reference to texts or other written materials).

4. Please take you unsigned exam to Karen she will process the documents in order to retain anonymity.

1. Short answer

You have been asked to investigate how *PT*, daily consumer expenditure on public transport in central London, is related to *C*, an index of its cost, and days of the week, classified as weekdays, Saturdays and Sundays. You have been given daily data for *PT* and *C* for the period May 1998–April 2003.

a) Explain, with a diagram, how you might specify a basic model that would reveal the differences in *PT* attributable to weekdays, Saturdays, and Sundays, stating the tests that you would perform.

b) It is generally thought that the sensitivity of *PT* to *C* is relatively low on weekdays, compared with Saturday and Sunday, because many of the weekday journeys are made by commuters who have no alternative to using public transport. Explain how you would test this proposition, providing a diagram to illustrate your specification.

c) Congestion pricing is an often proposed market mechanism to alleviate overuse of public goods such as roads. Briefly describe the economic rationale of such a pricing mechanism.

d) You are asked to investigate whether the Congestion Charge introduced in London February 2003 might have affected expenditure on public transport. Describe various ways of testing this. In particular, explain whether a Chow test might be suitable for this purpose.

e) At a seminar someone suggests that it might be legitimate to simplify the model by dropping the distinction between Saturdays and Sundays, combining them as weekend days. Explain how you might test this. (Assume that the proposition in (b) turned out to be correct.)

2. A researcher obtains a data set for a sample of 2,185 male respondents aged 35 to 42 in the US in the year 2000. It includes years of schooling, *S*, age, *AGE*, and hourly earnings in dollars, *EARNINGS*. Defining *LGEARN* as the natural logarithm of *EARNINGS*, the researcher regresses *LGEARN* on *S*, with the results shown in column (1) in the table (standard errors in parentheses; *RSS* = residual sum of squares). He realizes that work experience is also likely to be a determinant of earnings but thinks that there are no work experience data in the data set. Instead he defines a measure of potential work experience, *PWE*, as

$$PWE = AGE - S - 6$$

and regresses *LGEARN* on *S* and *PWE*, with the results shown in column (2) in the table.

As it happens, the data set did include actual work experience, *AWE*. If the researcher had regressed *LGEARN* on *S* and *AWE*, he would have obtained the results shown in column (3) of the table. The correlation between *S* and *PWE* was –0.74; the correlation between *S* and *AWE* was –0.28; and the correlation between *PWE* and *AWE* was 0.47. For the purposes of this question, it may be assumed that the third regression is the correct specification.

	(1)	(2)	(3)
S	0.1032 (0.0058)	0.1134 (0.0086)	0.1224 (0.0059)
PWE	–	0.0103 (0.0065)	–
AWE	–	–	0.0435 (0.0038)
constant	6.0131 (0.0787)	5.6700 (0.2305)	5.0603 (0.1125)
R^2	0.1274	0.1284	0.1776
Residual Sum Squares (RSS)	1048.0	1046.8	987.7

a) Explain intuitively and mathematically why the coefficient of *S* is smaller in column (1) than in column (3).

b) Given that the researcher did not know that the *AWE* data were available,

   i. Explain what benefits, in theory, he hoped to obtain by including *PWE* in the specification.

   ii. Evaluate whether the inclusion of *PWE* in the regression did improve the results in the way expected by the researcher.

c) The standard error of the coefficient of *S* is smaller in column (3) than in column (2). Explain whether this is what you would have expected.

d) The reduction in *RSS* on adding *PWE* to the specification in (2) is much smaller than the reduction when *AWE* is added in (3). Show numerically the relationship between the *RSS* in each equation and the respective *t* statistics for *PWE* and *AWE*. {recall the F-statistic formula in RSS form $F(j, n-k) = \frac{(RSS_R - RSS_{UR})/j}{RSS_{UR/(n-k)}}$ and the relationship be the F and t statistic when j=1}

e) Suppose that length of work experience were correlated with unobserved determinants of earnings such as motivation and responsibility. How would this affect the estimation of the coefficients in the third specification? (Give a general explanation; mathematical analysis is not expected.)

f) If length of work experience were correlated with unobserved determinants of earnings as in (e), suggest a suitable estimation procedure.

3. Limited dependent variable models (LDVM) have been an important part of the menu in applied econometrics for a number of years.

a) Explain in general terms what is meant by LDVM.
b) In this context explain what is the Linear Probability Model (LPM) using equations and graphs. Discuss the advantages and disadvantages of the LPM.
c) Discuss, using equations and graphs, alternatives to the LPM. Provide an empirical example to illustrate these models clearly explaining the research question to be examined, the data set required and the carefully define all variables. Explain how you would estimates these models.
d) Is a Tobit model an LDVM? Explain when the Tobit would be appropriate.

4.

a) Explain the difference between the use of a proxy variable and the use of an instrumental variable in a regression model.

b) A researcher has the following data for 40 cities in the United States for the year 2012: L, annual total purchases of lettuce from farmers markets per household, and P, the average price of a farmer's market lettuce in the city. She believes that the true relationship is

$$logL = \beta_1 + \beta_2 logP + \beta_3 logY + u$$

where Y is average household income, but she lacks data on Y and fits the regression (standard errors in parentheses):

$$\widehat{logL} = \underset{(0.52)}{13.74} + \underset{(0.23)}{0.17} logP \qquad R^2=0.01$$

Do you see anything nonsensical in this result? Explain analytically whether the slope coefficient is likely to be biased.

You are told that if the researcher had been able to obtain data on Y, her regression would have been

$$\widehat{logL} = \underset{(2.93)}{-1.63} - \underset{(0.21)}{0.48} logP + \underset{(0.35)}{1.83} logY \qquad R^2=0.44$$

You are also told that Y and P are positively correlated.

c) The researcher is not able to obtain data on Y but, from local authority records, she is able to obtain data on H, the average value of a house in each city, and she decides to use it as a proxy for Y. She fits the following regression (standard errors in parentheses):

$$\widehat{logL} = \underset{(3.22)}{-0.63} - \underset{(0.22)}{0.37} logP + \underset{(0.38)}{1.69} logY \qquad R^2=0.36$$

Describe the theoretical benefits from using H as a proxy for Y, discussing whether they appear to have been obtained in this example.

5. A variable Y_i, is generated as $Y_i = \beta_0 + \beta_1 X_i + u_i$. β_0 and β_1 are fixed parameters and u_i is a disturbance term that satisfies the Guass-Markov conditions.

You have the following sample data, recorded in the form of (Y_i, X_i):

(8, 6), (10, 2), (6, 4), (4, 2), (27, 21).

a) Calculate the Ordinary least squares (OLS) estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$. Make a simple graph of the data points and the estimated regression line.

Explain what is correct, incorrect, or incomplete in the following statements, giving a brief explanation if not correct.

b) Ordinary least squares (OLS) estimates of the parameters will be biased if the fifth observation is included in the sample because it is so far from the other observations. If it is dropped, the sample will be consistent, the OLS estimates of the parameters will be consistent, the standard errors will be smaller, the residual sum of squares will be smaller, and R^2 will be higher.

c) One way of dealing with the problem is to divide the specification through by X, fitting a model like

$$\frac{Y_i}{X_i} = \beta_0 \frac{1}{X_i} + \beta_1 + v_i$$

Where v_i is a disturbance term defined as $\frac{u_i}{X_i}$. By scaling through in this way, much less weight is given to the fifth observation.

d) Alternatively, given this sample of observations one could fill in the gap between X_4 and X_5 by construction new observations from the existing ones. For example, one new observation, say (Y_6, X_6) could be the average of the 5th and 4th observations as follows

$X_6 = (X_5+X_4)0.5$ and $Y_6 = (Y_5+Y_4)0.5$. One could continue to create observations by averaging the 3rd and the 5th, the 2nd and the 5th, and the 1st and the 5th. Besides filling gaps this will increase the number of observations and make the estimates of the parameters more accurate.

PHD Preliminary Examination – Econometrics – January 23, 2015 – 9am – 12 pm – Agricultural and Resource Economics

Choose 3 questions. Answer **all parts** of three (total) questions you choose.

You have a maximum of three hours. As your answers will be photocopied for grading, please make sure you press down when writing, leave a 1 inch margin on both sides, top and bottom, and write only on 1 side of the page. Please write down any assumptions you make as a part of an answer. For questions requiring an explanation, grading will depend on the explanation or rationale.

QUESTION 1.

A variable Y_i, is generated as $Y_i = \beta_0 + \beta_1 X_i + u_i$. β_0 and $\beta_1$are fixed parameters and u_i is a disturbance term that satisfies the Guass-Markov conditions.

You have the following sample data, recorded in the form of (Y_i, X_i):

(8, 6), (10, 2), (6, 4), (4, 2), (27, 21).

a) Calculate the Ordinary least squares (OLS) estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$. Make a simple graph of the data points and the estimated regression line.

Explain what is correct, incorrect, or incomplete in the following statements, giving a brief explanation if not correct.

b) Ordinary least squares (OLS) estimates of the parameters will be biased if the fifth observation is included in the sample because it is so far from the other observations. If it is dropped, the sample will be consistent, the OLS estimates of the parameters will be consistent, the standard errors will be smaller, the residual sum of squares will be smaller, and R^2 will be higher.

c) One way of dealing with the problem is to divide the specification through by X, fitting a model like

$$\frac{Y_i}{X_i} = \beta_0 \frac{1}{X_i} + \beta_1 + v_i$$

Where v_i is a disturbance term defined as $\frac{u_i}{X_i}$. By scaling through in this way, much less weight is given to the fifth observation.

d) Alternatively, given this sample of observations one could fill in the gap between X_4 and X_5 by construction new observations from the existing ones. For example, one new observation, say (Y_6, X_6) could be the average of the 5th and 4th observations as follows $X_6 = (X_5+X_4)0.5$ and $Y_6 = (Y_5+Y_4)0.5$. One could continue to create observations by averaging the 3rd and the 5th, the 2nd and the 5th, and the 1st and the 5th. Besides filling gaps this will increase the number of observations and make the estimates of the parameters more accurate.

QUESTION 2.

Selection bias has become an important consideration in various areas of applied micro-econometric analysis. One area where this issue has received particular attention is in labor economics.

(i) Please clearly define what is meant by selectivity bias and explain the econometric implications.
(ii) Provide an empirical example of your choosing and explain the source of the selectivity bias in your example.
(iii) Assuming you can obtain any data you might need, provide an empirical formulation of your model in general form and explain how you would test and correct for selectivity (if needed).
(iv) What is an exclusion restriction by definition and describe such a variable in your example.
(v) Are there any similarities between selectivity bias and endogeneity? Please illustrate with an example.
(vi.) You suspect a variable in your model is in fact endogenous and your office mate suggests using a proxy variable to avoid this endogeneity problem. Explain the difference between the use of a proxy variable and the use of an instrumental variable in a regression model. Should you follow this advice?

QUESTION 3.

Assume you are asked to provide estimates of the elasticity of demand for seasonal farm labor used in fruit farms in California. There are several econometric models that are plausible to address this question.

i. Please select one econometric model and then briefly explain its theoretical foundation.
ii. Clearly present the empirical model you would estimate, the data that would be required (assuming no budget restrictions on data collection) and the approach you would choose to select the sample and to collect the data.
iii. Provide the econometric specification of your model and explain how you would estimate it.
iv. Present a brief discussion of major weaknesses and strengths of your approach.
v. Finally, show how you would calculate the required elasticity.

(Question 4 appears on the next page)

QUESTION 4.

Assume that a consulting firm wants to hire you to undertake an econometric study designed to measure economies of size/scale in wheat farming in the United States. Before hiring can take place you must submit a proposal that carefully identifies two alternative research strategies that can provide sound measures of economies of scale. One specific objective of the study is to account for climatic variability and you need to incorporate this aspect in your proposal.

The Terms of Reference require that you have an explicit discussion of the following issues although additional consideration you deem relevant can be included:

i. A clear explanation of two alternative economic models using equations and graphs;
ii. The pros and cons of each of the two models;
iii. A definition of the specific measure of economies of size in each case;
iv. A specification of each model clearly defining all relevant variables;
v. A discussion of the data that would be needed and how it would be collected/obtained;
vi. The econometric procedures to be used in estimating each of the two models.

Ph.D. Preliminary Examination – Econometrics: June 6, 2014
Agricultural and Resource Economics
9am-1pm

Answer four of the five questions. You have a maximum of four hours. As your answers will be photocopied for grading, please make sure you press down when writing, leave 1 inch border on the side of the page, and write only on 1 side of the page. Also make sure your writing is legible or it will not be graded. If you make any assumptions as a part of an answer, please write down any assumption that you make. For questions requiring an explanation, points will be assigned for the explanation. Good Luck!!!!

1. The concept of mean square error (MSE) is a very important concept in statistics and econometrics as it brings together two important concepts: efficiency and bias.

a. Define efficiency and bias.

b. Define MSE.

c. Using a graph of the sampling distributions for two estimators of the same parameter and the concept of MSE, explain why a biased estimator may be preferred to an unbiased estimator. In your answer be sure to make use of the distinction between the estimator and the estimate.

2. A seed company hires you to determine the price they should charge farmers for their recently patented wheat variety.

a. Identify a list of variables you would need to determine the optimal price? Explain why you would need these variables and how you might go about obtaining the data.
b. Write down the economic model you would estimate to identify the optimal price and explain all aspects of the model.
c. Within the context of your model, do you anticipate problems with heteroscedacity, endogeneity, autocorrelation? Please explain how you would test for and address each problem if it were present.

3. Using monthly data from January 1990 to December 2000, an economic agent estimates the following single-equation model using OLS:

$$QDrP_t = 60 - 1.25PDrP_t + .40PCOKE_t + .30PPEPSI_t + .6INCOME_t$$
$$(3.0)\ (10.2) \qquad (5.1) \qquad (3.7) \qquad (4.9)$$

$R^2 = .90$ Durbin-Watson = .45

The number in parentheses corresponds to the t-statistic resulting from OLS estimation.

$QDrP_t$ = quantity of Dr. Pepper purchased per capita in gallons in month t
$PDrP_t$ = real price of Dr. Pepper in month t ($/gallon)
$PCOKE_t$ = real price of Coca-Cola in month t ($/gallon)
$PPEPSI_t$ = real price of Pepsi in month t ($/gallon)
$INCOME_t$ = real income per capita in month t in the US (1982-84 dollars)

Dr. Pepper, Coca-Cola, and Pepsi are carbonated soft drinks (soda).

(a) Assuming the estimation is correct, provide an economic interpretation of these results using appropriate terminology.

(b) Explain how you would provide a measure of the own-price, cross-price, and income elasticities of demand based on these results.

(c) Explain how you would test whether or not the elasticities obtained in part (b) are statistically different from zero.

(d) Are the properties for OLS to be the best linear unbiased estimator (BLUE) satisfied? Explain.

(e) Is it possible to improve on the econometric work provided by this economic agent? Explain your answer.

4. Dealing with heterogeneity in micro-econometric models has become an important issue in theoretical and applied work. Clearly define what is meant by heterogeneity and provide two well-articulated examples.

Then clearly show the econometric specification of the fixed and random effects models, explain when one would be preferable over the other and the key assumptions/implications of each.

Sketch how you would statistically test the fixed effects against the random effects model.

Finally, present an empirical example where you would need to apply panel data models clearly explaining the hypotheses to be investigated, the data needed, how such data would be obtained, the empirical equation(s) to be estimated, including a justification of the functional form chosen.

Choose an estimation strategy for your model and then justify this choice.

5. True, False or Uncertain. Explain your answer.

a. If two variables are independent their correlation coefficient will always be zero.

b. An unbiased estimator of a parameter (theta) means that it will always be equal to (theta)

c. An estimator can be BLUE (best linear unbiased estimator) only if its sampling distribution is normal.

d. OLS is an estimating procedure that minimizes the sum of errors squared.

e. Multiple linear regression is used to model annual income (y) using number of years of education (x1) and number of years employed in current job (x2). It is possible that the regression coefficient of x2 is positive in a simple linear regression but negative in a multiple regression.

f. Multiple linear regression is used to model annual income (y) using number of years of education (x1) and number of years employed in current job (x2). It is possible that R2 is equal to 0.30 in a simple linear regression model using only x1 and equal to 0.26 in a multiple regression model using both x1 and x2.