PhD/MA Econometrics Examination

August, 2017

Total Time: 8 hours

MA students are required to answer from A and B.

PhD students are required to answer from A, B, and C.

The answers should be presented in terms of equations, statistical details, and with necessary proofs and statistical deduction. Please use words, sentences, and paragraphs to describe your work. Equations alone will not be sufficient.

PART A
(Answer any TWO from Part A)

Q1. Fundamentals of OLS

a. Write out the OLS equation in matrix form. Also, write out the matrices and state their dimensions.
b. State the OLS assumptions in mathematical statements and in sentences (words).
c. Show that the OLS estimator is BLUE and define BLUE. Show all parts: B, L, U, and E.
d. What are the properties (hint: there are six) of the OLS estimator? State them in mathematics and words. Also, state any requirements which are necessary for these properties to hold.
e. Given the properties in part d, what can you infer about the disturbances from the residuals?
f. Write out a simple OLS model. Define your variables and describe how your model might meet or not meet all of the assumptions you stated above.

Q2. Probability Theory, Distributions, and More

a. State the probability axioms. Or define probability theory axiomatically.
b. What distribution is below:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

c. Find the first and second moments of this distribution directly from the distribution itself.
d. Find the first and second moments using the moment generating function. Also, discuss if these moments are the same or different from the moments you found in **part (c)** and why or why not.
e. This distribution has a property called “memoryless.” Prove that this distribution is memoryless.
f. Give definitions of the third and fourth moments and describe what those moments mean.
g. The distribution given in **part (d)** is a special case of another distribution. Name that other distribution.
h. Are there any other continuous distributions with the memoryless property?

Q3. The Variance of Least Squares
We know $var(b) = \sigma^2(X'X)^{-1}$, but σ^2 is an unknown parameter. Therefore in order to find $\widehat{var}(b)$, we need to find a good estimator for σ^2.

a. Derive that estimator.
b. Write down the standard error of b.
c. What is the standard of error of b used for? Or explain "why did we derive it?"

Part B: Answer any two of the following three questions

[Short verbal descriptive answer without mathematical proofs, steps, and necessary derivation will not earn you full credit.]

Q4. Suppose you want to examine the effect of a training program on the productivity of sales people in a firm. You decide to estimate the following model to measure the sales effect of training.

$$\log(sales) = \beta_0 + \beta_1 training + \varepsilon$$

where

training is a variable that indicates the number of hours in training

sales reflects the annual sales revenue of this individual

a. Which of the Ordinary Least Squares assumptions likely fails when estimating this model? Explain why? What does this mean for your estimate of the average effect of training on sales?
b. What is the "Fundamental Problem of Causal Inference"?

Suppose you decide to run an experiment intending to improve the productivity of sales people in the firm. There are two programs being tested, sales training *(training)* and a higher sales commission rate *(commission)*. For sales training, workers are randomly assigned into different number of hours each week (*training* is a variable for number of hours). For the new higher commission rate, workers are randomly placed into two groups, those who receive a higher commission and those who continue to receive the old commission rate (*commission* = 1 for people receiving the higher rate). The variable *sales*, again, reflects annual sales revenue of this individual and *experience* reflects years worked at the firm.

There are 204 sales peoples participating in this experiment

Table 1: OLS Regression Results

	Coefficients	
	Model 1	**Model 3**
Intercept	44.929149	44.929149
	(0.554151)	(0.554151)
Training	0.005000	0.010000
	(0.00100)	(0.001000)
Commission	0.020000	0.010000
	(0.00300)	(0.003000)
Experience	-0.001000	0.008000
	(0.002000)	(0.002000)
Experience X Commission		0.002000
		(0.000020)
Experience X Training		-0.002000
		(0.000020)
Degrees of Freedom	200	200
R-squared	0.1800	0.1900
Adjusted R-Squared	0.1795	0.1805

Standard Errors in Parentheses

Model 1:

$$\log(sales) = \beta_0 + \beta_1 training + \beta_2 commission + \beta_3 \exp erience + \varepsilon$$

You run the OLS regression and get the following results reported in the first column of **Table 1**.

c. What does the R-squared tell you about the regression?
d. Calculate the 95% confidence interval on $\hat{\beta}_1$. The critical value is 1.96
e. Interpret the coefficient estimate on training (*take for granted that the unbiasedness assumptions hold, which means randomization worked)*.

f. Interpret the coefficient estimate $\hat{\beta}_2$ from the model estimated (*take for granted that the unbiasedness assumptions hold, which means randomization worked)*.
g. Suppose you wanted to test the hypothesis that neither intervention had any effect on sales, clearly state the null hypothesis you would test.

You estimate the following model

Model 2:

$$\log(sales) = \beta_0 + \beta_1 \exp erience + \varepsilon$$

and find a sum of squatted residuals of 420. Test the hypothesis that neither intervention affects sales (the critical value at the 1% significance level for this problem is 4.506). Clearly walk through the steps to test this hypothesis, and state the conclusion of the test.

h. Suppose the F-statistic you calculated in part (g) has a p-value of 0.015. What does this p-value tell you about the null hypothesis.

A co-worker suggests the following estimated model better explains sales.

Model 3

$$\log(sales) = \beta_0 + \beta_1 training + \beta_2 commission + \beta_3 \exp erience +$$
$$\beta_4 (\exp erience * commission) + \beta_5 (\exp erience * training) + \varepsilon$$

You run the OLS regression and get the results reported in Column 2 of **Table 1.**

i. Derive an interpretation for the marginal effect of an additional hour of training. (Take for granted the OLS assumptions hold, and that our coefficients' estimates are significant).
j. Explain how experience and job training interact in this sales department.
k. Derive an interpretation for how receiving the higher commission rate affects sales.
l. Do more experienced or less experienced sales reps respond better to the new commission incentive? (Which coefficient and what about it answers this question).
m. Suppose a coworker points out that training likely helps sales a lot at low levels, but the returns to training decrease (and perhaps even turn negative) as reps receive more and more of it. What variable would you add to Model 1 to estimate this effect and test this hypothesis?

Q5. Suppose $\tilde{y}$ is an unobserved latent variable that measures an individual's economic productivity (which can be proxied by hourly earnings), such that:

$$\tilde{y} = x\beta + \varepsilon \text{ where } \varepsilon \sim N(0, \sigma^2 I)$$

However, you do not observe earnings in your data. You only observe, y_i, which indicates whether an individual is working or not. $y_i = 1$ if an individual participates in the labor force and $y_i = 0$, otherwise. An individual participates in the labor force if he/she is able to earn wages above some reservation wage, w.

Define $\phi(\theta)$ as the pdf for a standard normal and $\Phi(\theta)$ as the cdf for the standard normal.

Note: $\frac{\partial \Phi(z)}{\partial \theta} = \phi(z) \frac{\partial z}{\theta}$

a. Define y_i in terms of $\tilde{y}_i$ and w.

b. What is θ, the identifiable parameter of interest in this problem?

c. Derive the probabilities that $y_i = 1$ and $y_i = 0$ for individual *i.*

d. *D*erive the contribution of each individual in your sample to the overall likelihood function (i.e., derive $\mathcal{L}_i(\theta)$) and the individual log-likelihood function.

e. Derive the score function needed to identify $\hat{\theta}_{MLE}$.

Now suppose you observe the earnings of each individual *only if* he/she participates in the labor market, such that $y_i = \tilde{y}_i$ if the individual participates in the labor market but is unobserved otherwise.

Note: $\frac{\partial \phi(z_i)}{\partial \theta} = -z_i \phi(z_i) \frac{\partial z_i}{\partial \theta}$

f. Now redefine y_i in terms of $\tilde{y}_i$ and w under this new setup.

g. Derive the probability that you observe each individual *i.* Assume $w = 0$, from this point forward.

h. *D*erive the contribution of each individual in your sample to the overall likelihood function (i.e., derive $\mathfrak{L}_i(\theta)$) and the individual log-likelihood function.

i. Derive the score function needed to identify $\hat{\theta}_{MLE}$.

j. Explain what is implied by the simplified form of the Score function (i.e., what is the implied orthogonality condition).

Q6. Consider the following model $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, where X_1 is a matrix of k_1 variables and X_2 is a matrix of k_2 variables such that

$$X_1 = \begin{bmatrix} x_{11}^1 & x_{11}^2 & \cdots & x_{11}^{k_1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n}^1 & x_{1n}^2 & \cdots & x_{1n}^{k_1} \end{bmatrix}, \; X_2 = \begin{bmatrix} x_{21}^1 & x_{21}^2 & \cdots & x_{21}^{k_1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{2n}^1 & x_{2n}^2 & \cdots & x_{2n}^{k_1} \end{bmatrix}.$$

Denote b_1 and b_2 as the Ordinary Least Squares estimates for β_1 and β_2, respectively.

a. Derive the expression for the ordinary least squares estimator b_1 as a function of Y, X_1, X_2, and b_2 using the partitioned regression model.

b. Suppose you only observe X_1 but not X_2. Thus you run the OLS model $Y = X_1\beta_1 + \varepsilon$.

   i. Derive the expression for OLS estimate of b_1 that you would estimate under these conditions (i.e., what is the usual OLS estimator for b_1 when you regress Y on X_1).

   ii. If $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ is the true model, give an expression for the amount b_1 (that you estimated in b.i is biased in this circumstance as a function of X_1, X_2, and b_2.

c. Now suppose you observe both X_1 and X_2. Derive the ordinary least squares estimator for and b_2 as a function of Y, X_1, X_2, and M_1 using the partitioned regression model. Where M_1 is the residual-maker matrix and $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$.

d. Define the Frisch-Waugh Theorem and describe its intuition.

Under what conditions is the bias you solved for in part b.ii equal to zero. What does this mean in the context of the Frisch-Waugh Theorem (i.e., what happens when you regression X_2 on

PART C: Answer any Two

[Short verbal descriptive answer without mathematical proofs, steps, and necessary derivation will not earn you full credit.]

Q7. Nepal Study Center is planning to study to help a clinic in a rural village in Nepal's Gulmi District to implement a micro health insurance program. It plans to use a dichotomous choice experiment design to carry out the study. The plan is to sample 420 households randomly from the three communities that lay around the clinic --its catchment area. Each community has nine wards. The sampling will be performed by using the proportional sampling design representing all the wards from each of the clinic. The households are presented with options to enroll in one of three micro health insurance plans: Basic (clinic visits), General (clinic + plus pharmacy), Comprehensive (clinic visits, pharmacy + minor surgery). The three alternatives are presented below:

c=(Comprehensive, General, Basic)

We would expect a person's utility related to each of the three alternatives to be a function of both personal characteristics (such as income, age etc..) and characteristics of the health care plan (such as its price/premium).

We collected data would look like the table below: person's age (divided by 10), the person's household income (in Rs10,00 / month), and the price of a plan (in Rs100 / 6 months). The first three cases from the data are shown below. It is in the long form.

	HHid	MH_Alt ch	Choice	hhinc	age	Premium
1	1	Comprehensive	1	3.66	2.1	2
2	1	General	0	3.66	2.1	1
3	1	Basic	0	3.66	2.1	0.5
4	2	Comprehensive	0	3.75	4.2	2
5	2	General	1	3.75	4.2	1
6	2	Basic	0	3.75	4.2	0.5
7	3	Comprehensive	0	2.32	2.4	2
8	3	General	0	2.32	2.4	1
9	3	Basic	1	2.32	2.4	0.5

Additionally, we will also collect information on the following variables: **Receive Remittance (yes/no), No Of Children, No of Clinic Visits Per Six Month, and Distance to Clinic (minutes of walking distance).** These variables are not shown in the table to save space.

Taking the first case (**id==1**), we see that the case-specific variables **hhinc, age, Remittance, NoChildren, and Distance** are constant across alternatives, whereas the alternative-specific variable **price** varies over alternatives. Additionally, we also collected information on the following variables: **Receive Remittance, No Of Children, No of Clinic Visits Per Six Month, and Distance to Clinic**

The variable **MHalt** (micro health insurance alternatives) labels the alternatives, and the binary variable **choice** indicates the chosen alternative (it is coded 1 for the chosen plan, and 0 otherwise).

Q. For simplicity, consider only three variables for model set up (age, income, and price).

A) Set up a Random Utility Model (RUM). Show all the steps (e.g., present indirect utility functions.)

B) Present the corresponding data table in the wide form as a set up for a long-hand mle coding.

C) Present the log likelihood function. Show all the steps.

(You may assume that the income and age have the same impact on the choice functions.)

D) As you recall, clogit automatically suppresses alternative specific constants, whereas asclogit allows the constants. In the DC modeling community, there is no consensus regarding the preference for an ASC (alternate specific constant approach versus the non-ASC option). In this case, which option may make more sense and why?

Q8. Given the following 2-equation system,

$$ChildLabor_t = \alpha_0 + \alpha_1 * MotherEduc_t + \alpha_2 * ChildAge_t + \alpha_3 * FatherRemittance_t + u_t$$

$$FatherRemittance_t = \beta_0 + \beta_1 * Poverty_t + \beta_2 * FatherAge_t + v_t$$

FatherRemittance variable captures the absence of fathers working overseas sending money back home to the family.

a. How would you estimate the *child labor* equation using a 2-sls method? Briefly discuss the two steps.
b. Likewise, how would you estimate the *Remittance* equation? Explain the necessary steps.
c. Set up this system in a grand matrix notation

$$Y = Z * \beta + U \quad (1)$$

d. Derive the Var-Cov(U). You can use the generic notation.
e. Discuss the iterative 3-SLS estimation method, complete with the necessary steps and derivations.

Q9.

Q 9.1 Consider the following probability density function:

$$f(y_t) = \frac{1}{\beta} * \ e^{-y_t/\beta}$$

Where *y* takes on positive values.

a. Set up the likelihood log likelihood function and obtain the *mle* estimator of the underlying parameter *β.*
b. Explain the method of calculating the variance of the estimator of *β.*
c. Also find the method of moment estimator of *β*.

OR

Q. 9.2 Consider the following epidemiological model for the State of Texas, where the children's asthma rate (y – proportional to the children's population from 1-6 years of age) was expressed as a function of PM2.5 airborne pollution count (x). The data were collected for 254 counties for the year of 2010.

$$y = \frac{\emptyset \, x}{\delta + x} + \text{e}$$

which can be generically expressed as: $\text{y} = f(x, \emptyset, \delta) + e$

For simplicity, the subscript "t" is suppressed.

a. Using the generic expression 2, present the numerical estimation algorithm using the Newton Raphson optimization algorithm OR the maximum likelihood estimation method (*e* can be assumed to follow a normal distribution).
b. Discuss the method of deriving the variance-covariance matrix of the estimators.
c. The parameter $\emptyset$ is also known as the maximum y, thus it cannot be more than 1 (i.e., 100% asthma rate). With that in mind, can you conjecture some sort of graphical relation between y and x for the model given above?