PhD/MA Econometrics Examination

January 2020

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. Verbal and brief descriptive discussions will not suffice.

PART A

(Answer any TWO from Part A)

1. ***Probability Theory, Distributions, and More***
    a. State Bayes' Theorem
    b. In words, describe what Bayes' Theorem means
    c. What distribution is below:

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

    d. Find the first and second moments of this distribution directly from the distribution itself.
    e. 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 d** and why or why not.
    f. This distribution has a property called "memoryless." Prove that this distribution is memoryless.
    g. Name a practical application of this distribution, i.e., where does it 'naturally' occur?
    h. Name the third and fourth moments – *name them do not calculate them.*
    i. Define and draw the PDF and CDF of the distribution.

2. *Ordinary Least Squares (OLS)*

- **a.** State the classical assumptions – in words and equations.
- **b.** Derive the normal equations.
- **c.** Demonstrate that the OLS estimator is BLUE.
- **d.** From $X'e = 0$, we can derive several properties. State these properties. Hint: there are 6 and 5 of them require that the OLS regression includes a constant.
- **e.** Write out a simple OLS model. Define your variables and describe how your model might meet or not meet the assumptions you stated above.

Question 3

Table 1

Source	SS	df	MS		Number of obs	
					F(*blank*)	
Model	1.31E+12	4	3.28E+11		Prob > F	0.000
Residual	1.48E+13	12397	1.19E+09		R-squared	
Total	1.61E+13	12401	1.30E+09		Root MSE	34523

Wage Income	Coef.	Std.Err.	t	P>t	[95% Conf.	Interval]
Age	3358.85	140.11	23.97	0.00	3084.22	3633.49
Age^2	-35.59	1.74	-20.41	0.00	-39.01	-32.17
Family Size	3208.56	504.99	*blank*	*blank*	*blank*	*blank*
Family Size^2	-442.89	58.92	-7.52	0.00	-558.38	-327.39
Constant	-50889.86	2691.38	-18.91	0.00	-56165.38	-45614.34

‘*blank*’ is intentional.

Table 2

Source	SS	df	MS		Number of obs	******
					F(5, 12396)	******
Model	1.59E+12	5	3.17E+11		Prob > F	0.000
Residual	1.45E+13	12396	1.17E+09		R-squared	******
Total	1.61E+13	12401	1.30E+09		Root MSE	34203

Wage Income	Coef.	Std.Err.	t	P>t	[95% Conf.	Interval]
Age	2825.81	143.12	19.74	0.000	2545.27	3106.34
Age^2	-31.09	1.75	-17.74	0.000	-34.53	-27.65
Family Size	-263.60	549.42	******	******	******	******
Family Size^2	-138.37	61.68	-2.24	0.025	-259.27	-17.47
Married	******	751.98	15.29	0.000	10025.89	12973.88
Constant	-36431.60	2829.10	-12.88	0.000	-41977.07	-30886.13

3. *Regression Output Interpretation, Calculations, Definitions, and More*

REFERRING ONLY TO TABLE 2 (parts a – f)

Use mathematical statements and words in your answers.

a. Fill in the missing values ***??????*** in Table 2. (There are 8 numbers to calculate.)
b. Define and interpret R-squared
c. Explain the relationship between F and R-squared
d. Define and interpret the 95% CI for "Family Size"
e. Interpret the coefficient on "Married"
f. Calculate the marginal effect of being one year older than the mean age (40.66272) of the sample.

COMPARING TABLES 1 & 2 (part g)

g. Why does adding "Married" create a relatively large change on the coefficients on Family Size and Family Size^2 but cause relatively small change the coefficients on Age and Age^2?

MORE (part h)

h. If you were going to add a variable to this regression, what variable would you add? Explain the variable choice (why?), what the expected sign would be, and how you interpret the coefficient on your newly added variable.

Did you calculate the missing numbers (??????**) in Table 2 (part a)?*

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.]

4. Consider the following demand equation for vanilla

$$Q_i^d = \beta_0 + \beta_1 P_i + u_i \qquad (1)$$

where Q_i^d is baker $i's$ demand for vanilla, P_i is the price, and u_i represents the error term, which also includes demand shifters.

a. If you were to regress (1) using Ordinary Least Squares (OLS), which of the OLS assumptions is likely to fail? Explain why? What does this mean for your estimate for the slope of your demand equation?

b. Suppose the supply equation for vanilla is as follows:

$$Q_i^s = \alpha_0 + \alpha_1 P_i + v \qquad (2)$$

where Q_i^s is vanilla supply, P_i is the price, and v_i represents the error term.

c. Assuming market equilibrium, using equations (1) and (2), derive an expression for P_i as a function of the slope parameters and error terms. How does this expression relate to the failed OLS assumption you named in (a).

d. Derive an expression for $cov(P_i, u_i)$ as a function of the slope parameters and the variance of u_i. What is the sign of $cov(P_i, u_i)$? *Hint: substitute the expression you derived in (c.) for p. Another Hint:* $cov(aX + bY, X) = avar(X) + bcov(X, Y)$.

e. Derive an expression for $cov(P_i, Q_i^d)$ in terms of β_1, $var(P_i)$, and $cov(P_i, u_i)$. Hint: use equation (1) to write Q_i^d as a function of P_i and substitute that function in for Q_i^d in $cov(P_i, Q_i^d)$.

f. The probability limit of the OLS estimate forβ_1 is as follows: $plim(\widehat{\beta_1^{OLS}}) = \frac{cov(P_i, Q_i^d)}{var(P_i)}$. Using the expression you derived in (e), calculate the asymptotic bias of $plim(\widehat{\beta_1^{OLS}})$ (i.e., what is $plim(\widehat{\beta_1^{OLS}}) - \beta_1$). What is the sign of that bias?

g. Madagascar is one of the primary vanilla producers in the world. Suppose you decided to instrument P_i in equation 1, using cyclone exposure in Madagascar. What conditions would this instrument have to satisfy? Do you think this instrument is a valid instrument for vanilla price? Explain why or why not.

h. Assume that Madagascar cyclone exposure is a valid instrument for P_i in equation (1). Denote cyclone exposure as Z. Using the variable names in in equation (1) and Z, describe the two-stage least squares regression process. Clearly define your first and second stages. Derive an expression for β_1^{IV} based on these two stages.

5. Let $\tilde{y}$ be some unobserved latent variable such that

$$y_i = \tilde{y}_i = X_i{}'\beta + \varepsilon_i \text{ if } \tilde{y}_i > 0$$

$$y_i \text{ is unobserved otherwise}$$

$$\text{and } \varepsilon \sim N(0, \sigma^2 I)$$

Note: $\dfrac{\partial \Phi(z)}{\partial \theta} = \phi(z)\dfrac{\partial z}{\theta}$ and $\dfrac{\partial \phi(z_i)}{\partial \theta} = -z_i \phi(z_i)\dfrac{\partial z_i}{\partial \theta}$

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

b. Derive the probability that you observe an individual *i.*

c. Derive the contribution of each individual in your sample to the overall likelihood function (i.e., derive $L_i(\theta)$) and the individual log-likelihood function. *(5 points)*

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

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

6. 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 egression 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.

PART C: 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.]

Q7.

a. Compare and contrast briefly the following estimation methods – OLS, GLS, Method of Moment (MM) & Generalized MM. No derivation is required.

b. Under what practical circumstances (modeling issue) would you use an exponential form: $f(y) = \gamma e^{-\gamma y}$ versus the Poisson form: $f(y) = e^{-\mu}\mu^{y}/y!$? You may present a research problem as an example.

c. Pick one of the pdfs –exponential or Poisson, and derive its mle and MM estimates.

d. Assume the following model: Y = XB + E => E = Y - XB, which can be transformed by multiplying it by a set of instruments Z'. This may lead to a quadratic form of minimization problem $ArgMin_{\widehat{\beta}}||\, Z'\hat{E}||$.

Carry out the GMM method to derive GMM estimator, sometimes known as efficient GMM. Under what condition, this estimator reduces to an IV estimator?

Q8. 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.

$$y = \frac{\emptyset\, x}{\delta + x} + e \quad (1)$$

which can be generically expressed as: $y = f(x, \emptyset, \delta) + e$ (2)

For simplicity, the subscript "t" is suppressed.

a. Using the generic expression 2, present the numerical estimation algorithm using the Newton Raphson optimization method. You can use NR method in either of two cases:

   a) Non-linear least squares method,
   OR
   b) The maximum likelihood estimation method, where the error term *e* can be assumed to follow a normal distribution.

b. Now, let's assume that you calculate the marginal impact of the pollution on the asthma rate --$^{dy}/_{dx} = f(x, \emptyset, \delta), complex\ non - linear\ expression.$ Explain a method of choice to calculate the (standard error) confidence level.

Q9. Nepal Study Center is planning to conduct a 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).

1.1. 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.

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.)

1.2. Note: cogit 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?

1.3. Usually, we use some sort of clustering adjustment for the standard errors vce (cluster id). In this case, which clustering id would you use – individual id, community id, or ward id – and why?