PhD/MA Econometrics Examination

January, 2015

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.

PART A

(Answer any TWO from Part A)

Question 1

Probability Theory and Distributions

a. What distribution is below:

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

b. Find the first and second moments of this distribution directly from the distribution itself.
c. 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 b** and why or why not.
d. This distribution is one of two distributions with a property called "memoryless." Prove that this distribution is memoryless.
e. Name a practical application of this distribution, i.e., where does it naturally occur?
f. The distribution above is a special case of another distribution. What distribution is it a special case of and exactly how is it a special case of that other distribution?

Question 2

Data Problems

- **a.** Multicollinearity
  - **i.** Define multicollinearity and how do you detect it?
  - **ii.** Describe a situation with perfect multicollinearity. How would address it?
  - **iii.** Describe a situation with near-perfect multicollinearity. How would address it?
- **b.** Measurement Error
  - **i.** Dependent variable – describe in words and equations
  - **ii.** Independent variable – describe in words and equations
  - **iii.** Which of the two, dependent or independent variable measurement error, is easier to deal with? Describe (in detail) why it is easier to deal with and how you would deal with the easier one.
- **c.** Missing Observations
  - **i.** There are four commonly used techniques used to deal with missing observations. Name and describe (in detail) three of the four techniques. Discuss each technique in terms of benefits, problems, implementation, etc.

Question 3

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 a number of properties. State these properties. Hint: there are 6 and 5 of them require that the OLS regression includes a constant.

PART B

(Answer any TWO from Part B)

Question 4

Consider the following linear model in matrix notation:

$$Y = X\beta + U \qquad (1)$$

And assume that the Var-Cov(U) = Ω (not a white noise residual.)

a. Using the decomposition method, derive the GLS estimator of β.
b. Derive the variance-covariance matrix of $\widehat{\beta_{gls}}$
c. Demonstrate that under the OLS assumptions of white noise, the GLS formulas in **Q. 1 a** and **Q. 1 b** reduce to the OLS formulas.
d. Assuming that the $Var(u_t) = \sigma_t^2 = \gamma z_t^2$(not a constant variance), present the matrices Ω and its decomposed component, P. (Assume that there is no serial autocorrelation.)
e. Now consider a simple 2 parameter regression equation – y(t) = b0 + b1 *x(t) + u(t). Using the transformation matrix P derived in d, transform this simple regression model and derive the weighted least squares (WLS) regression.
f. When would you use robust standard error?
g. How would you test for heteroscedasticity?

Question 5

Consider the following scenario where consumer confidence is assumed to be affected by the expected oil price fluctuation.

$$ConsumerConfidence_t = \alpha_0 + \alpha_1 \, OilPrice_t^* + u_t$$

Where, * on Price variable means expected price.

a. Set up the adaptive expectations mechanism.
b. Derive the Koyck formulation.
c. Discuss the estimation method. That is, will the application of OLS be valid? Why or why not? Provide a solution.

Question 6

Imagine we know with certainty that the following model fully describes the true state of the supply and demand for wine. First, the demand for wine in any year, ***Q(t)***, is a function of the price of wine, ***pw(t);*** the income of the average individual, ***I(t);*** and the price of beer, ***pb(t)***.

Second, in any year the price of wine is a function of the amount of wine brought to market, ***Q(t)***, and a weather index, **W=*DroughtIndex.*** Third, the error terms in the supply and demand functions are due purely to measurement errors—that is, there are no omitted variables in the model. Thus, we have the following two equation model:

Demand: $Q(t) = \alpha 0 + \alpha 1\ pw(t) + \alpha 2\ I(t) + \alpha 3\ pb(t) + \varepsilon(t)$ ***(1) and***
Supply: $pw(t) = \beta 0 + \beta 1\ Q(t) + \beta 2\ W(t) + \eta(t)$ ***(2)***

We assume that the error terms each are normally distributed with a mean of zero and a constant variance. Moreover, we assume that the two error terms are independent of each other—that is, we are assuming that: $E(t) \sim N(0, \sigma 2\_\varepsilon)$, $\eta(t) \sim N(0, \sigma 2\_\eta)$.

Finally, we assume that income, the price of beer, and the weather index are non-stochastic variables—i.e., these variables are independent of the two error terms.

a. Present the endogenous variables. List the predetermined variables.
b. Identify the demand and the supply equations using the order condition.
c. Present the 2 SLS method of estimation. Under this method of estimation, which of the two assumptions are you following: a) $E(\varepsilon(t).\eta(t)) = 0$. Or b) $E(\varepsilon(t).\eta(t)) \neq 0$, *and why?*
d. Using the supply equation as an example, discuss the generated regressor bias problem and show how you would correct for it.
e. How would you test if the ***Q(t)*** variable in the supply equation is endogenous?
f. Using the supply equation as an example, describe the method of testing for the over-identifying restrictions.

PART C

(Answer any TWO from Part C)

Question 7

Consider the following VAR-model:

$$InfUS(t) \quad = C1 + \ \phi 11 * InfUS(t-1) \quad + \phi 12 * InfMexico(t-1) + u1(t)$$
$$InfMexico(t) \ = C2 + \ \phi 21 * InfUS(t-1) \quad + \phi 22 * InfMexico(t-1) + u2(\text{t})$$

a. Derive the variance-covariance matrix of the two error vectors: u1(t) and u2(t).
b. Present the FGLS method of estimating this VAR model.
c. Why is it called a seemingly unrelated regression (SUR), and not a simultaneous model?
d. Under what condition, would the GLS method be equal to the OLS?
e. How would you test the causal effect from Mexico to US?
f. Given the fact that these two regressions have lagged variables on the right hand side, is there any other way you could detect the contemporaneous relationship between the two country's inflation? Explain.

Question 8

A research project is studying the level of lead in home drinking water as a function of the age of a house, family income, and the city blocks. Assume that there are 5 city blocks where they live in. The water testing kit cannot detect lead concentrations below 5 parts per billion (ppb), and it also cannot register above 15 ppb. That is, the data collected will be censored from both ends.

a. Set up this model with proper notation and equations.
b. Set up the log likelihood function for this model.
c. Describe the likelihood ratio test procedure to test the effect of geographic location on the lead content in the water. (set up the null, alternate, etc..)
d. How would you calculate the (unconditional) mean of the level of lead in the water?

Question 9

Assume that you are interested in modeling the drinking behavior of the young men between the age of 18 and 35 years. The survey was conducted and the question was asked as follows: "Over the last 6 weeks, how many drinks have you had?" Response ranged from 0 to 62. The data was collected on variables such as, gender, age, education, and income. There are several count modeling approaches: Poisson, Negative Binomial, Hurdle, and even Zero Inflated Poisson.

a. First, describe each approach carefully emphasizing the scenario under which each of the method would be appropriate.
b. Then, propose the best method that would be appropriate for the scenario described above.
c. Set up the log-likelihood function for your proposed method in b.
d. Discuss the appropriate algorithm to optimize the log-likehood function.
e. How would you calculate the marginal impact of income on the number of drinks?
f. How would calculate the marginal impact of gender on the number of drinks?