ECON 21020: Midterm Examination

Allotted Time: 1 Hour and 20 Minutes

Problem 1

(20 Points.) A climate-distressed farmer wants to learn about the methane gas emissions from her cows. She collects an *i.i.d.* sample $\{X_i\}_{i=1}^n$, where X_i is the amount of methane (in pounds) emitted from a cow in a single day. She assumes $X_i \sim \text{Uniform}[0, \kappa]$, i.e. X_i has the pdf:

$$f_{X_i}(x) = \begin{cases} 1/\kappa, & \text{if } 0 \leq x \leq \kappa \\ 0, & \text{otherwise} \end{cases}$$

(a) Show that $E(X_i) = \kappa/2$ and $\text{Var}(X_i) = \kappa^2/12$.

(b) Given the sample $\{X_i\}_{i=1}^n$, write down a method of moments estimator for $\theta = 1/\kappa$.

(c) Is your estimator from (b) unbiased? If so, prove it. If not, specify the direction of the bias.

Problem 2

(30 Points.) Consider the model $Y = \beta X^2 + U$. You make the assumption that $E(Y|X) = \beta X^2$.

(a) Show that $E(U|X) = 0$.

(b) Are X^2 and U correlated? *Hint: the correlation coefficient is* $\rho_{X^2,U} = \frac{\text{Cov}(X^2,U)}{\sqrt{\text{Var}(X^2)\text{Var}(U)}}$

(c) Show that $\beta^* = E(X^2Y)/E(X^4)$ is the b that minimizes $\text{MSE}(b) = E[(Y - bX^2)^2]$.

(d) Given an *i.i.d.* sample $\{Y_i, X_i\}_{i=1}^n$, write down the method of moments estimator $\hat{\beta}_n$ for β^*.

(e) Argue that $\hat{\beta}_n$ is both consistent and unbiased for β^*. *Hint:* $E(\hat{\beta}_n) = E(E(\hat{\beta}_n|\{X_i\}_{i=1}^n))$.

Problem 3

(30 points.) Suppose you are interested in predicting students' future incomes based on whether or not they go to private high schools. You write down the model $Y = \beta_0 + \beta_1 X + U$, where:

- Y is someone's annual income in adulthood (in dollars)
- X is an indicator for attending a private high school ($1 =$ *private*, $0 =$ *public*)

Part I

Assume $\beta_0 + \beta_1 X$ is the *best linear predictor* of Y given X, so that $E(U) = E(UX) = 0$. Given an *i.i.d.* sample $\{Y_i, X_i\}_{i=1}^n$, with variation in X_i, you compute the OLS estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

(a) State whether or not $\hat{\beta}_0$ and $\hat{\beta}_1$ satisfy each of the following properties. If the property is not satisfied, list what additional assumption(s) you would need for the property to hold.

- (i) $\hat{\beta}_0$ and $\hat{\beta}_1$ are consistent.
- (ii) $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased.
- (iii) $\hat{\beta}_0$ and $\hat{\beta}_1$ are the best linear unbiased estimators *(BLUE)*.

(b) Suppose that you compute $\hat{\beta}_1$ and $\text{se}(\hat{\beta}_1)$. Explain how you would test whether β_1 is significantly different from zero at the 5% level (i.e. $\alpha = 0.05$). That is, state the appropriate test statistic T_n and critical value c_n. Then, write down a 95% confidence interval for β_1. *Hint: you may assume that n is large, and you do NOT need to derive the test statistic.*

Part II

Now you want to give β_1 a causal interpretation. Therefore, you interpret β_1 as the *effect* of attending a private high school on someone's future income. Your goal is to estimate β_1.

(c) Suppose $E(U) \neq 0$. Show that the model can always be re-written with the same slope, but a new intercept and error term, so that the new error has an expected value of zero.

(d) Is U likely to be correlated with X? Briefly explain your reasoning.

(e) Will running OLS uncover the causal effect of X on Y? Briefly explain.

Problem 4

(20 points.) Consider the linear model $wage = \beta_0 + \beta_1 age + \beta_2 educ + \beta_3 child + U$, where *wage* denotes earnings, *age* is how old someone is, *educ* represents years of education, and *child* denotes number of children. Assume that your best linear predictor (BLP) assumptions hold.

(a) Give a causal interpretation to β_1.

(b) Suppose you do not observe *educ* and *child*, so you consider the linear model with just *age*:

$$wage = \gamma_0 + \gamma_1 age + \varepsilon$$

Assume that the BLP assumptions hold for this simple linear model: $E(\varepsilon) = E([age]\varepsilon) = 0$. What assumptions would you need to make in order to guarantee that γ_1 equals β_1?

(c) Suppose you wish to test whether age^2 and age^3 are jointly relevant in the original model. In other words, you want to test whether the coefficients on age^2 and age^3 would both equal zero if they were included in the original regression. Describe which linear regressions to run and how to construct the test given an *i.i.d.* sample $\{wage_i, age_i, educ_i, child_i\}_{i=1}^n$.