

Ph.D. Econometrics I
Heinz School, Carnegie Mellon University
90-906, Spring 2004

Midterm

Instructions You may use any books, notes, calculators, and other aids you like. You may not converse, nor may you cooperate.

Please complete all questions.

Please show all relevant work.

Please interpret your results in plain English.

Use the backs of pages if you need more space.

Each part of each question worth 10 points, except for 4a, which is worth 20.

1. Consider the following regression model in which we assume that all the classical assumptions are satisfied:

$$Y = \beta_1 X + \epsilon \quad (1)$$

Consider the following estimator for β_1 :

$$\tilde{\beta}_1 = \frac{\sum X_i^2 Y_i}{\sum X_i^3}$$

- (a) Is $\tilde{\beta}_1$ unbiased?

(b) What is the variance of $\tilde{\beta}_1$?

(c) Is the variance of the OLS estimator greater than, less than, or the same as the variance of $\tilde{\beta}_1$?

2. Consider a factory producing pairs of shoes. We have data on the monthly costs of running the factory and the number of shoes produced. We want to estimate the following model by OLS:

$$\text{Cost}_i = \beta_1 + \beta_2 \text{Left Shoes} + \beta_3 \text{Right Shoes} + \epsilon$$

Is it a good plan to try to estimate this model by OLS? Why or why not?

3. Suppose you are interested in trying to predict how much money people will spend on their next new car. You have a dataset documenting how much a large, relevant sample of people spent on their cars and how much they characteristically spend on a bottle of wine. So, you estimate the following model by OLS:

$$P_{\text{car}} = \beta_1 + \beta_2 P_{\text{wine}} + \epsilon$$

Then, for a similar group of people, you are informed of each person's P_{wine} , and you use your estimates of β_1 and β_2 to predict car purchase price, P_{car} .

A critic complains that the model above does not conform to the assumptions of the classical linear regression model. Is the critic right about this?

The critic goes on to claim that, since the assumptions of CLRM are violated, you should not use OLS for these purposes. Assuming that the critic is right about the assumption violation, is she right about the uselessness of OLS here?

4. The Medical Expenditure Panel Survey is an annual survey which collects information about medical expenditures, income, employment, demographics, health information, &c for a representative sample of Americans.

I have prepared an extract of these data for 1996. The following variables appear on the data:

Variable	Meaning
age	age of person in years
sex	sex of person, 1=male & 0=female
income	income in 1996 \$
employed	1=employed, 0=not employed
insured	1=had health insurance, 0=not
health	perceived health status, higher is sicker
spending	spending on health care, 1996 \$

- (a) Consider the regression on page 5 of the output. Please interpret each of the estimates there, excluding the constant. Which ones “make sense” intuitively?

(b) For which of the coefficients can we reject, at the 5% level, the null hypothesis that the coefficient is equal to zero?

- (c) Suppose an unemployed and uninsured person took a job with health insurance. What is your best estimate of the change in their health spending? Can we conclude at the 10% level that this change is different from 0?

(d) If we were to regress spending on income alone, would the slope coefficient be positive or negative and why?